

```

US-09-985-689A-3 (1-433) x TA31SH10P (1-574)
QY 18 GlyLeuThrGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 63 GGTATGACGGTAGTGGTGGAGATATAGGGGTAGCGGATACGGGTATCGACTTT-----116
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuThrAlaLeuGly 57
Db 117 AACAGCTGTTTCTTCCAGATCCA-----AATCAAGAGGTGGCGCTTTTCCCGAAGTT 170
QY 58 ArgThrAsnAsn-----61
Db 171 AACTATACACCGCAAAATCGTGCATTTGCCCGTGTGACTTCATCCGGGGGATATAC 230
QY 62 -----AlaAsnAspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db 231 TTTGTGGGGATGAGGAAATAGTGTGATGCGACGACGTCGGCAGGTACCGCAGCGGGAGT 290
QY 76 ValLeu-----GlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 291 GTATTATGTAACACGGTAAATGCCAAGTATAATGTGTGCGCAAGGGGGCGNAGATTTC 350
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 351 TTCAGGGGTGTGGGTGCCATCCAGTCAGAGCTTGTCTCCCGCCACGCGTTACTCAA 410
QY 113 LeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 132
Db 411 ATTATTCGTCGGATATGGCGTGGAGCCGCTGTGTCTCAACTCGGGGTTTGT 470
QY 133 ValAsnGlyAlaThrThrAlaAsnSerArgGlnValAspGluThrValArgAsn---Asn 151
Db 471 GTCCTCCCGAGTATCTGCTGGAAAGGATATGATGAGTTTGGCAGTAGTTATGAC 530
QY 152 AspMetThrValLeuPheAlaAlaGlyAsn 161
Db 531 GATGGCTACTTATCTTCCACTGGCAAC 560

RESULT 12
AJ273402 614 bp mRNA linear EST 29-DEC-1999
LOCUS AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
DEFINITION cDNA clone Ma#948, mRNA sequence.
ACCESSION AJ273402.1 GI:6432774
VERSION AJ273402.1
KEYWORDS EST.
SOURCE Metarhizium anisopliae
ORGANISM Metarhizium anisopliae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipitaceae; Clavicipitaceae; Metarhizium.
REFERENCE 1 (bases 1 to 614)
AUTHORS Screen,S.E., Mathur,P. and St. Leger,R.J.
TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae
JOURNAL Unpublished (1999)
COMMENT Contact: Screen SE
Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA.
FEATURES
Source 1..614
/mol_type="mRNA"
/strain="ARSEF 2575"
/db_xref="taxon:5530"
/clone="Ma#948"
/note="Vector: UniZap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector, UniZap"
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1,09e-08 Length: 614
Score: 188.00 Matches: 68
Percent Similarity: 45.68% Conservative: 43
Best Local Similarity: 27.98% Mismatches: 84
Query Match: 8.31% Indels: 48
Db: 9 Gaps: 12

US-09-985-689A-3 (1-433) x AJ273402 (1-614)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 11 GGTGAGGTACTTGCCTATATATATATATATGACACTGGTATTGAG-----52
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuThrAlaLeuGlyArgThrAsn 60
Db 53 GCTCCACACCCAGTTTGGGTGCGGCACCTTTCTTAAGAGCTTCATCAGCGTCAA 112
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 113 AAC---ACTGATGCCACGCCCATGGGACTCACTGGCTGGTACCATT-----GGTAGC 163
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer---99
Db 164 AAGACCTACGGTGTGCCAAAAGGCTAAGCTATGCTGTCAGGTTCTTGACACACAG 223
QY 100 ---SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 224 GGCAAGTGGTCTACTCCGCTATCATCAGTGGCATGCTACCTGTCACAGGACTCCAAG 283
QY 119 AsnAlaGly-----AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGly 135
Db 284 ACCGCGCGCTGCCCAACGGCGCATTTGCTTCCATGATGAGCTGGGA-----GGT 331
QY 136 AlaThrThrAlaAsnSerArgGlnValAspGluThrValArgAsnAsnAspMetThrVal 155
Db 332 GGCTACTCGCGCTCCCAACAGGCTGCTGCTTGGTCAATCTGCTGCTTCCCT 391
QY 156 LeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAla 175
Db 392 GCGTCTCGCGCTGCAACGATACCGGATGCGCCAGAACACCTCT---CCGCTTCGAG 448
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnThrArgProSerPheGlySerIleAla 195
Db 449 CCTTCTGCTGCACTGTTGGTGTCTCT-----GCG 478
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 479 GAAATGACAGCGCATCTCTCTTCTCACTAC-----GCGAGAGTT 520
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 521 ---GTCGATATTTTCGCTCTGCTGAGCAATGTTCTTCCACC-----559
QY 236 AspSerSerPheTrpAlaAsnThrAsnSerLysThrAlaThrMetGlyGlyThrSerMet 255
Db 560 -----TGG-----ATTGGTGGCGGCAACACCATCTCTGTGACTCTCATG 601
QY 256 AlaThrPro 258
Db 602 GCTACTCTCC 610

RESULT 13
BZ424995/c 2121 bp DNA linear GSS 13-DEC-2002
LOCUS BZ424995 Aspergillus terreus random genomic DNA clone library
DEFINITION 10023066-5726 Aspergillus terreus genomic, genomic survey sequence.
ACCESSION BZ424995
VERSION BZ424995.1 GI:26666450
KEYWORDS GSS.
SOURCE Aspergillus terreus
ORGANISM Aspergillus terreus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

```

REFERENCE 1 (bases 1 to 2121)
 AUTHORS Akenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J., Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and Madden,K.T.
 TITLE Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing strains
 JOURNAL Unpublished (2002)
 COMMENT Contact: Zimmer DP
 Microbia, Inc.
 One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
 Tel: 617-621-8322
 Fax: 617-621-8322
 Email: dzimmer@microbia.com
 Class: plasmid ends.
 Location/Qualifiers
 1. 2121
 /organism="Aspergillus terreus"
 /mol_type="genomic DNA"
 /strain="ATCC 20542 (A. terreus Thom, anamorph)"
 /db_xref="taxon:33178"
 /lab_host="Escherichia coli"
 /clone_lib="Aspergillus terreus random genomic DNA clone library"
 /note="vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI; Sau3A genomic fragments ligated into BamHI digested pZEROTM-2"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.66e-07 Length: 2121
 Score: 185.50 Matches: 79
 Percent Similarity: 38.97% Conservative: 34
 Best Local Similarity: 27.24% Mismatches: 86
 Query Match: 8.20% Indels: 92
 Gaps: 12
 DB: 28
 US-09-985-689a-3 (1-433) x B2424995 (1-2121)
 QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 1281 GCGAGGACACTAGCTAGCTGTAGACACCGCATCACGTGGAC-----1234
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
 DB 1233 -----CACGAGGAATTTGAGGGCGTGGAGTCTCGGCTACACGGCTGCGGGGCGAG 1180
 QY 61 AsnAlaAsnApproAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 DB 1179 CATGTG---GACGGGTGCGGCATGGCCACCCACCGTTCTGTACAAATGGTGGTAAGACA 1123
 QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle-MetAspSerSe 100
 DB 1122 TAC-----GGGTGGCCAGAGGCCAACCTGCTG-----TCGGTCAAGGTCTTCGTC 1075
 QY 100 tGlyGlyLeuGlyLeuProSerAsnLeuAsn-----ThrLe 113
 DB 1074 GGGGAATCGAGCAGCAGCATTCATCTCTGGACGGCTTCAACGGCTTCAACGGCTTCTGTC 1015
 QY 113 uPheSerGln-----AlaTrpAsnAlaGlyAlaAr 123
 DB 1014 GTTTCAGAGCGGCACCTGGAAAGCGGAGCATCAACATGAGCTTGGGTATGTTATTCTGC 955
 QY 123 gIleHisThrAsnSerTrp-----GlyAlaProValAsnGlyAlaTy 137
 DB 954 CATCTATGTGATGA-TATCATTTGGAGTCACAGCTGACGCAAGCATAGCGGTGGATA 896
 QY 137 rThrAlaAsnSerArgGlnValAspGlyTyValArgAsnAsnAspMetThrValLeuPh 157
 DB 895 CTCCAAGGCTTCATATGATCGCGTCGAGACCGCTTCAACGGGAGGTCCTGTCATCGT 836
 QY 157 eAlaAlaGlyAsnGluGly-----ProAsnSerGlyThrIleSer-- 170

Db 835 CCCCCCGGCAATGAGATGTGGTGCACGCTTCCACCTGTGATGGAGTGGGACTCC 776
 QY 171 -----AlaProGly-ThrAlaLysAlaIai 179
 Db 775 GTGACTGATTCGACAGACGCGCTCGAGCAGCGCGGCTTCTGCTCTGATGCC 716
 QY 179 leThrValGlyAla-----ThrGluAsnTyArgProSerPheGlySerI 194
 Db 715 TTACCGTGGCTGCGATCAACGTGAACAACACCGTGCCTATTTCTCCAACTACGGCTCG 656
 QY 194 leAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyA 214
 Db 655 TGGTG-----651
 QY 214 tGileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuA 234
 Db 650 -----GATATCTTCGCGCGCGGCGGAGACATCTCTCTGCC-----615
 QY 234 laProAspSerSerPheTrpAlaAsnTyArgAsnSerLysTyAlaTyMetGlyGlyThrS 254
 Db 614 -----TGGATCGGCTCCAAACACGCGCCACCAACCATCTCGGACCT 572
 QY 254 erMetAlaThrProIleValAlaGly 262
 Db 571 CCATGCGCCACCCCCACATTTGTGCGC 546
 RESULT 14
 BQ770462 665 bp mRNA linear EST 26-JUL-2002
 UI-M-F10-byv-m-19-0-UI-r1 NIH_BMAP_F10 Mus musculus cDNA clone
 IMAGE:5702970 5', mRNA sequence.
 BQ770462
 BQ770462.1 GI:21978936
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.nhl.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pYX-5.
 Location/Qualifiers
 1. 665
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5702970"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP_F10"
 /note="Organ: Brain; Vector: pYX-Asc; Site1: EcoR I;
 Site 2: Not I; The library was constructed according
 to Bontal, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated

FEATURES
source

with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCACGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:
 Pred. No.: 4,59e-08 Length: 665
 Score: 182.50 Matches: 65
 Percent Similarity: 42.68% Conservative: 37
 Best Local Similarity: 27.20% Mismatches: 96
 Query Match: 8.06% Indels: 41
 DB: 13 Gaps: 7

US-09-985-689A-3 (1-433) x BQ770462 (1-665)

```

QY 63 AsnAppProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
DB 14 GATGATGGGTAGGCATGGCCATCTCGTTCGAGGTGTATGCCAGCATGAGGAGTGC 73
QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly 101
DB 74 CAAGGATTTGCTCCAGATGCAGAGCTGCACATCTTCAGGGTCTTTACCAACAATCAG--- 130
QY 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrPheAsnAlaGly 121
DB 131 -----GTGCTTACACATCTTCGTTCTGGATGCTTCAACTATGCC 172
QY 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsnSer 141
DB 173 ATCTAAAGATGAGACGCTTCTCAACCTAGCATCGTGGCCCGCATCTCATGATCAT 232
QY 142 ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla 159
DB 233 CGGTTTGTTCACAAAGGTGGGAATTAACAGCATCAATGTAATATGTTCTCTATT 292
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAsnAlaIle 179
DB 293 GGCATGATGAGACCTCTCTATGGCACTCTGAATAACCTCTGATCATGATGATGAT 352
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn 199
DB 353 CGAGTGGGTGGC-----ATTGACTTTGAAGAT 379
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGly 213
DB 380 AACATCGCTCGCTTTCTTCCAGGGGAATGACTACTCTGGGAATACCAGGAGCTATG 439
QY 214 ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerLeu 233
DB 440 CGTGTGAAGCTGCACATTCACCTATGCTGCTGGAGTGGGGGT----- 484
QY 234 AlaProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr 253
DB 485 -----TCCGGTGTGAAGGGGGCTGGCTGCATCTCAGGGACC 523
QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 273
DB 524 AGTGTGCTTCNCCAGTGGTGGTGGGGCGCTCACCTTGTAGTAAAGCAGATACAGAAG 583
QY 274 AsnArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAla 292
DB 584 CGGAGAGCTGGT-----AATCTGCCAGTGTGAAGCAGCTTTGATGCGTCAGCC 634

```

RESULT 15
 BZ8933395/c 716 bp DNA linear GSS 30-JUL-2003
 LOCUS HL2_0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
 DEFINITION genomic survey sequence.

ACCESSION BZ8933395
 VERSION BZ8933395.1 GI:33343985
 KEYWORDS GSS.
 SOURCE Halorubrum lacusprofundi
 ORGANISM Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Halobacteriaceae; Halorubrum.

REFERENCE 1 (bases 1 to 716)
 AUTHORS Goo Y., Reach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M., Dasarma, S., Ng, W.V. and Hood, L.
 TITLE Low-pass Sequencing for Microbial Comparative Genomics
 JOURNAL Unpublished (2003)
 COMMENT Contact: Goo Y
 Institute for Systems Biology
 1441 North 34th Street, Seattle, WA 98103, USA
 Tel: 206 732 1412
 Fax: 206 732 1299
 Email: ygoosystemsbiology.org
 Seq primer: M13 Forward
 Class: shotgun.

FEATURES
 source
 1..716
 /organism="Halorubrum lacusprofundi"
 /mol_type="genomic DNA"
 /strain="ATCC 49239"
 /db_xref="taxon:2247"
 /clone_lib="H1 pUC18 Library"
 /notes="Vector: pUC18; Site_1: SmaI; A shotgun library was constructed from Halorubrum lacusprofundi genomic DNA using pUC18/SmaI/BAP plasmid"

ORIGIN

Alignment Scores:
 Pred. No.: 5,21e-08 Length: 716
 Score: 182.50 Matches: 71
 Percent Similarity: 45.45% Conservative: 34
 Best Local Similarity: 30.74% Mismatches: 66
 Query Match: 8.06% Indels: 61
 DB: 28 Gaps: 11

US-09-985-689A-3 (1-433) x BZ8933395 (1-716)

```

QY 60 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu---Gly 78
DB 701 TCCGATGCTCGGACGACGAGCGGACGGAATCACGTCCGCGAACCGTCCGGCGGA 642
QY 79 AsnAlaLeuAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 96
DB 641 AACGCGAGTGGGACCGCAATCGCGCTCGCACCGACGCGTCCACGGGATCAAGGTG 582
QY 97 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerG 116
DB 581 TTCGACGACGACGGA-----CCACGCGACGCTTCGTCGCGCTCGTCGCG 537
QY 116 nAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaProva 133
DB 536 GGAATGGAACACGCGACGAGGATCCGACGTGCTTCAGATGACCTCGGTGCG 477
QY 133 LAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMe 153
DB 476 GACGGGCACTTACATACTTATCATCGAACCG-GTT---CGGAACACTCGCAGTCCGGGAA 421
QY 153 tThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProG 173
DB 420 GATCGCGTCTGTTTCGGCGGGAACATCGGT-----CAAGGAACGTCGAGTTCCTCCCG 367
QY 173 yThrAlaLeuAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySe 193
DB 366 GAACGCTTACACTCGCTCGGTCGGAGCGGTTCACGACGCGC----- 321
QY 193 rIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThr----- 210
DB 320 -----GGCGTCCGCGCACTTCTCCAGC---GGAGAGACGATCAACAC 283

```

```

Qy 211 -----ArgAspGlyArgIleLeuProAs 218
Db 282 GTCAGCGCGTGGGAAAGTATGCCCCCGCGACTGGCCCGACGAGTACGTGGTTCGGA 223
Qy 218 pValThrAlaProGlyThrPheIleLeuSerAlaArg-----SerSerLeuAlaPr 235
Db 222 CGTGTGCGCGCCAGCGGTAGCGTATACTCGGCGGACCGGCGGTACAAACATCCGAAA 163
Qy 235 oAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMe 255
Db 162 GGAC-----GGCACCTCCAT 148
Qy 255 tAlaThrProIleValAlaGlyAsnValAla 265
Db 147 GGCCGCACCGCACCGTACGCGCGCTCGCGGCG 117

```

Search completed: April 4, 2004, 11:54:47
Job time : 2284.58 secs


```

Qy 119 AsnAlaGly-----AlaArgileHisThrAsnSerTrpGlyAlaProValaAsnGly 135
Db 284 ACCCGGGCTGCCCAACAGCGGCATTCCTCCATGAGCTGGGA-----GGT 331

Qy 136 AlaTyrThrAlaAsnSerArgGlnValaAspGluTyrValaArgAsnAspMetThrVal 155
Db 332 GGCTACTCGCGCGTCCGCAACCAAGAGTGGCTGTGTGTCATTCCTGCTTCCTT 391

Qy 156 LeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAla 175
Db 392 GCCGTGCGCGCTGCCACAGATACCGGGATGCCAGACACTCT---CCCGTTCGGAG 448

Qy 176 LysAlaAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAla 195
Db 449 CCTTCGCTCCGCTGCTGTGTGCTCT---GGC 478

Qy 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgile 215
Db 479 GAAATGACAGCGCATCTCTCTTCACTAC-----GGCAGAGTT 520

Qy 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 521 ---GTCGATATTTTCGCTCCTGTAGCAATGTTCTTCCACC-----559

Qy 236 AspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMet 255
Db 560 -----TGG-----ATTGGTGGCGGCACAAACCAACCACTCTCTGTACTCNCATG 601

Qy 256 AlaThrPro 258
Db 602 GCTACTCCC 610

RESULT 13
TA315H10P 574 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 315h10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL490202
VERSION AL490202.1 GI:11866292
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 574)
AUTHORS Hall N., Bowman S., Lennard N.J., Doggett J., Atkin R.,
Chillingworth C., Ormond D., Harris B., El-Sayed N., Hou L.,
Melville S.E., Rajandream M.A. and Barrell B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: neilsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
Location/Qualifiers
1..574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="315h10"

ORIGIN

```

```

Alignment Scores:
Pred. No.: 9.18e-09 Length: 574
Score: 168.00 Matches: 53
Percent Similarity: 47.65% Conservative: 28
Best Local Similarity: 31.18% Mismatches: 59
Query Match: 8.32% Indels: 30
DB: 29 Gaps: 7

US-09-985-689A-4 (1-433) x TA315H10P (1-574)
Qy 18 GlyLeuTyrGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 63 GGTATTGACGGTAGTGGTGAGATATAGCGGTAGCCGATACGGTATCGATTT-----116

Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 117 AACAGCTGTTTCTTCCAGCATCCA-----AATCAAGAGGTGGCGCTTTACCCGAAGTT 170

Qy 58 ArgThrAsnAsn-----61
Db 171 AACTATAACCCCGCAAAATCGTGTCAATTCGCCCGTGTGACTTCATCCGGGGGATTAC 230

Qy 62 ---AlaSerAspProAsn---GlyHisGlyThrHisVal-----AlaGlySer 75
Db 231 TTTGCTGGGATGAGGAATAGGTTCATGCGACGCGCAGGTGCGAGGTACCGCAGCGGGAGT 290

Qy 76 ValLeu-----GlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 291 GTTATTAGTAACGCGGTAAATGCCAAGTAAATGTTGTCGCAAGGGGGCGAAGATTTC 350

Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 351 TTCAGGGGGTGGGTGGCCATCCAGTCAGAGCTGTCTCTCCCCACGACTTACTCAA 410

Qy 113 LeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 132
Db 411 ATTATTCTCCCGGATATGCGCTGAGCCCGTGTGTCTCAAACTCGTGGGGTTTGT 470

Qy 133 ValaAsnGlyAlaTyrThrAlaAsnSerArgGlnValaAspGluTyrValaArgAsn---Asn 151
Db 471 GTCCTCTCGAGTATTCGCTGTGGAAGGATATGGATGGATTCGAGTAGTTATGAC 530

Qy 152 AspMetThrValLeuPheAlaAlaGlyAsn 161
Db 531 GATCGCTACTTATCTCTCCACTGGCAAC 560

RESULT 14
BZ893395/c 716 bp DNA linear GSS 30-JUL-2003
LOCUS HL2_0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
DEFINITION genomic survey sequence.
ACCESSION BZ893395
VERSION BZ893395.1 GI:33343985
KEYWORDS GSS.
SOURCE Halorubrum lacusprofundi
ORGANISM Halorubrum lacusprofundi
REFERENCE 1 (bases 1 to 716)
AUTHORS Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
Goo Y., Roach J., Glusman G., Baliga N.S., Deutsch K., Pan M.,
Dassarma S., Ng W. V. and Hood L.
Low-pass Sequencing for Microbial Comparative Genomics
Unpublished (2003)
Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers

```

```
source
1..716
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="HL pUC18 Library"
/notes="Vector: pUC18; Site 1: SmaI; A shotgun library was constructed from Halorubrum lacusprofundi genomic DNA using pUC18/SmaI/BAP plasmid"
ORIGIN
Alignment Scores:
Pred. No.: 2,428-08 Length: 716
Score: 185.50 Matches: 72
Percent Similarity: 45.45% Conservative: 33
Best Local Similarity: 31.17% Mismatches: 66
Query Match: 8.21% Indels: 61
DB: 28 Gaps: 11
US-09-985-689A-4 (1-433) x BZ893395 (1-716)
QY 60 AsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu---Gly 78
Db 701 TCCGATGCGTCGACGACGAGCGGACGGAACCTACGTCGCGGAAACCGTCGCGGCGGA 642
QY 79 AsnAlaLeuAsnLys-----GlyMetAlaProGlnAlaAsnLeuAlaPheGlnSerIle 96
Db 641 AACGGAGTGGGACCGCAATCGCGCTCGCACCGAAGCGTGTCTCCACGGATCAAGGTG 582
QY 97 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerG 116
Db 581 TTCGACGACGACGGGA-----CCAAACGGGAGCGTTCGTCGCGGTGTCGCGC 537
QY 116 nAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaProVa 133
Db 536 GGAATGGAACACCGGACGCGAGCTCGGACGTCGACGTCCTTCAGATGAGCCTCGTCGC 477
QY 133 lAsnGlyAlaTrpThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMe 153
Db 476 GACGGGCACTTACCATCTTCATCGAACCG-GTT---CGGAACACTCGAGTCGCGGAA 421
QY 153 tThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProG 173
Db 420 GATCGCGGTGCTTTCGCGCGGGAACATCGGT-----CAAGGAACGTCGAGTCTCCCGG 367
QY 173 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySe 193
Db 366 GAACGTCACGACTCGCTCGCGTGGAGCGGTCAACGACACGCCG----- 321
QY 193 rIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThr----- 210
Db 320 -----GGCGTCGCGCACTTCTCCAGC---GGAGACACGATCAACAC 283
QY 211 -----ArgAspGlyArgIleLysProAs 218
Db 282 GTCAAGCGCGTGGGAGTATGATCCCGCGGACTGCGCGGACGAGTACGTGTTCCGGA 223
QY 218 pValThrAlaProGlyThrPheIleLeuSerAlaArg-----SerSerIleAlaPr 235
Db 222 CGTGTGCGGCGCCAGGCGTCGAGCGTATACTCGGCGGAACCGGCGGTACAAACATCCGAA 163
QY 235 oAspSerSerPheTrpAlaAsnTyrAsnSerIysTyrAlaTyrMetGlyGlyThrSerMe 255
Db 162 GGAC-----GGCACCTCCAT 148
QY 255 tAlaThrProIleValAlaGlyAsnValAla 265
Db 147 GGCGGACCGCACGTCAGCGCGCTCGCGGG 117
RESULT 15
BQ770462 665 bp mRNA linear EST 26-JUL-2002
LOCUS BQ770462
DEFINITION UI-M-F10-byy-m-19-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
```

```
IMAGE:5702970 5', mRNA sequence.
BQ770462
VERSION BQ770462.1 GI:21978936
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbsr@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
```

Seq primer: pYX-5.

```
FEATURES
source
1..665
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702970"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_F10"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CACCACGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
```

ORIGIN

```
Alignment Scores:
Pred. No.: 5,5e-08 Length: 665
Score: 181.50 Matches: 65
Percent Similarity: 42.44% Conservative: 36
Best Local Similarity: 27.31% Mismatches: 96
Query Match: 8.03% Indels: 41
DB: 13 Gaps: 7
```

US-09-985-689A-4 (1-433) x BQ770462 (1-665)

```
QY 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLys 83
Db 17 GATGGGTAGGCCATGCGACATCGTTCGAGGTGTGATCCGACGATGAGGAGTGCCAA 76
QY 84 GlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGlyGly 102
Db 77 GGATTTCCTCCAGATGCGAGAGCTGCACATCTTCAGGTCTTTTACCACCAATCAG----- 130
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 131 -----GTGCTTACACATCTGTGTTCTCGATGCTTCACTGATCCATGCGCATC 175
```

```
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 176 CTAAGAAGATGGACGTTCTCAACCTTAGCATCGTGGCCGCCACTTCATGGATCATCCG 235
Qy 143 GlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
Db 236 TTTCCTTGACAAGGTGTGGGAATTAACACCTAACAAATGTAATATGTTCTGCTATTGGC 295
Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 296 AATGATGGACCTCTCTATGGCACTCTGAATAACCTCTGATCAGATGGATGGATTGGA 355
Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
Db 356 GTGGGTGGC-----ATTGACTTTGAAGATAAC 382
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGlyArg 214
Db 383 ATCGCTCGCTTTCTTCAGGGGAATGACTACCTGGGAATTACCAGGAGGCTATGGTCGT 442
Qy 215 IleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 234
Db 443 GTGAAGCCTGACATGTACCTATGCTGCTGGAGTGGGGGT----- 484
Qy 235 ProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSer 254
Db 485 -----TCCGGTGTGAAGGGGGCTGCCGTGCACCTCTCAGGACCCAGT 526
Qy 255 MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsn 274
Db 527 GTCGGCTTCNCAGTGGTGGTGGGGCGCTACCTTGTTAGTACACAGTACAGACCGG 586
Qy 275 ArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAla 292
Db 587 GAGCTGGTG-----AATCCTGCCAGTGTGAAGCACGCTTTGTAGCGTCAGCC 634
```

Search completed: April 4, 2004, 11:54:55
Job time : 2282.58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3544.27 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVPSGPQPSLAIVH 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh

-Q=/cgr2_1/USPTO.spool/US09985689/runat_31032004_161806_4155/app_query.fasta_1.3498
-DB=GenEmbl -QFMT=fastop -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -ILOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cui -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CCN 1 1 11925 @runat 31032004 161806 4155 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.nam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2261	100.0	1299	1	AB046405 Bacillus
2	2251	99.6	1299	1	AB046402 Bacillus
3	2245	99.3	2218	6	E03808 DNA encodin
4	2240	99.1	1299	1	AB046404 Bacillus
5	2044	90.4	3003	6	AR069954 Sequence
6	2044	90.4	3003	6	BD062155 Nucleic a
7	2024.5	89.5	1302	1	AB046406 Bacillus
8	2003.5	88.6	1923	1	AB084155 Bacillus
9	2002.5	88.6	1920	1	AB046403 Bacillus
10	2002.5	88.6	1920	6	AR368116 Sequence
11	1999.5	88.4	1923	6	AR368118 Sequence
12	1998.5	88.4	1305	6	AX839476 Sequence
13	1998.5	88.4	1923	1	AB051423 Bacillus
14	1998.5	88.4	1923	6	AR368117 Sequence
15	612.5	27.1	60006	1	AF268611 Unculture
16	558.5	24.7	301205	1	AE017218 Geobacter
17	501	22.2	5890	3	UU60086 Dictyosteli
18	465	20.6	5772	3	AF466309 Dictyoste
19	463.5	20.5	6115	3	DDU20432 Dictyostell
20	443	19.6	1977	6	AR201152 Sequence
21	441.5	19.5	5109	3	AF263455 Dictyoste
22	422.5	18.7	1236	6	AR201146 Sequence
23	422.5	18.7	1962	6	AR201155 Sequence
24	422.5	18.7	12452	1	AE010265 Pyrococcu
25	402.5	17.8	14376	1	AE013049 Thermoana
26	399	17.6	126928	2	AC096673 Trypanoso
27	395.5	17.5	303450	1	SCO939130 Streptomy
28	377.5	16.7	299850	1	AP004601 Oceanobac
29	370.5	15.5	2539	6	AR202322 Sequence
30	347.5	15.4	3348	1	D83872 Sequence
31	345.5	15.3	1239	1	AF305633 Thermoana
32	345.5	15.3	1239	1	AY028704 Thermoana
33	345.5	15.3	2809	6	AR202321 Sequence
34	345.5	15.3	3900	1	AB007809 Streptomy
35	344	15.2	292550	1	AP001513 Bacillus
36	343.5	15.2	12295	1	AE013026 Thermoana
37	341.5	15.1	135638	1	AF484556 Streptomy
38	338	14.9	300800	1	SCO939112 Streptomy
39	334	14.8	1329	6	AX433519 Sequence
40	329	14.6	302300	1	AF005034 Streptomy
41	328	14.5	2731	1	AF015225 Thermooc
42	328	14.5	6854	1	BSU39230 Bacillus sp
43	328	14.5	291000	1	SCO939105 Streptomy
44	327	14.5	300425	1	AP005044 Streptomy
45	324	14.3	301150	1	AP004602 Oceanobac

ALIGNMENTS

RESULT 1

```

AB046405
LOCUS       AB046405             1299 bp    DNA    linear    BCT 23-JAN-2001
DEFINITION   Bacillus sp. SD521 PROD gene for protease, partial cds.
ACCESSION   AB046405
VERSION     AB046405.1   GI:12381942
SOURCE      Bacillus sp. SD521
ORGANISM    Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
REFERENCE   1 (sites)
AUTHORS     Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
            Horikoshi, K.
TITLE       Novel oxidatively stable subtilisin-like serine proteases from
            alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
            evolutionary relationships
JOURNAL     Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE     20568675
PUBMED      11118284
REFERENCE   2 (bases 1 to 1299)
AUTHORS     Saeki, K.
TITLE       Direct Submission
JOURNAL     Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
            Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
            Tochigi 321-3497, Japan (E-mail: 387185@kcastanet.kao.co.jp,
            Tel: 81-285-68-7400, Fax: 81-285-68-7403)
FEATURES             Location/Qualifiers
     source          1..1299
                     /organism="Bacillus sp. SD521"
                     /mol_type="genomic DNA"
                     /strain="SD521"
                     /db_xref="taxon:133780"
     gene            1..1299
                     /gene="PROD"
     CDS             1..1299
                     /gene="PROD"
                     /codon_start=1
                     /transl_table=11
                     /product="protease"
                     /protein_id="BAB21268.1"
                     /db_xref="GI:12381943"
                     /translation="NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRDSSNHE
                     AFRGKITALYALGNTNANDPNFGHTVAGSVGNALNGKAPQANLVFQSIIDSSGG
                     LGGLPSNLNTLFSQAMNAGARIHTNSWAGPANGAYTANSQVDEYVYRNDMTVLFAAG
                     NEGNSTGTSIAPGTAKNAITVGATENYRPSFGSLADPNPHIAQFSRGRATRDGRKPD
                     VTAPGTITLARGAPDSGSFWANYNSKYAYMGTSMTATPIVAGNVAQLREHFINKRG
                     ITPKPSLTKAALAGATDVLGVPSPQGWGRVTLDKSLNAVYNEATLATLQKATY
                     SFOQAQKPKISVWTDAPGSTASYTLVNDLQNLVTPNGOKYVGNDESYPYDNNW
                     DGRNNVENFINAPQSGTYTIEVQYNNVPSGPRFSLAIVH"
ORIGIN
Alignment Scores:
Pred. No.:      5,48e-134      Length:      1299
Score:          2261.00      Matches:      433
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             1              Gaps:        0

US-09-985-689a-5 (1-433) x AB046405 (1-1299)
QY      1  AenaspValAlaAatgGlyileVallyseAlaAaspValAlaGlnAenAenTyrGlyLeuTyr  20
DB      1  AATGATGAGCAAGAGGAATAGTAAACGACGTTGACCAAAACAATTACGGACTATAT  60
QY      21  GlyGlnGlyValValAlaValAlaAaspThrGlyLeuAaspThrGlyArgAenAaspSer  40
DB      61  GGACAAGGTCAGTAGTTGCGTAGCGGATACCGGTTAGATACAGTCGTACGATAGT  120
QY      41  SerMetHisGluAlaPhaArgGlyLysileThrAlateuTyrAlaLeuGlyArgThrAsn  60
DB      121  TCTATGCAATGAAGCAATCCGGGGAAATACAGCTCTTTACGCGTAGGAAGCACTAAC  180
QY      61  AenAlaAenAaspProAenGlyHisGlyThrHisValAlaGlySerValLeuGlyAenAla  80
181  AATGCAATGATCCGAATGGGCATGTGTCACGATGTAGCAGTCTCTGTACTTGGTAATGCT  240
81  LeuAenlysglyMetAlaProGlnAlaAenLeuValPheGlnSerileMetAaspSerSer  100
241  TTAATAAAGGATGGTCCGCAAGCTAACTTAGTCTTCCAACTATTATGGATAGCAGC  300
101  GlyGlyLeuGlyLeuProSerAenLeuAenThrLeuPheSerGlnAlaTrpAenAla  120
301  GGAGGATAGTGGATACCATCGAATTCGAATACGTTATTAGTCAAGCTTGAATGCT  360
121  GlyAlaArgileHisThrAenSerTrpGlyAlaProValAenGlnAlaTyrThrAlaAen  140
361  GGGCTAGATTCACTTAATCTTGGGGTCTCCAGTAATAGGAGGTACACTGCTAAC  420
141  SerArgGlnValAaspGluTyrValArgAenAenAaspMetThrValLeuPheAlaGly  160
421  TCGAGACAAGTGGATGAGTATGTTGCAATAATGATGATGACGGTACTTTTTGACGAGGT  480
161  AenGlnGlyProAenSerGlyThrIleSerAlaProGlyThrAlaLysAenAlaIleThr  180
481  AATGAAGTCTTAATTCAGGAACAATTAGTGCTCCAGGCACAGCGAATAAATGCCATTACG  540
181  ValGlyAlaThrGlnAenTyrArgProSerPheGlySerLeuAlaAaspAenProAenHis  200
541  GTCCGGCGCAACGAAACTATCGCCGAGCTTCGGTTCAATTAGCAGATAACCCAAATCAT  600
201  IleAlaGlnPheSerSerArgGlyAlaThrArgAaspGlyArgileLysProAaspValThr  220
601  ATCGCACAAATTTTCATCAAGAGGAGCTACGAGATGGACGAATTAACCCAGACGTACAC  660
221  AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAaspSerSerPheTrp  240
661  GCTCTCGAACAATTTATTATCAGCAGCTTCTTCTTAGCCCCAGACTCTTCGTTTGG  720
241  AlaAenTyrAenSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal  260
721  GCGAATTAACAGTAAGTATGCGTACATGGCGGTACCTCTATGGGACACCTATAGTT  780
261  AlaGlyAenValAlaGlnLeuArgGluHisPheIleLysAenArgGlyIleThrProLys  280
781  GCGGGGATGTCGCGCATATTACGCGAGCATTTTATAAAAATAGAGGAATTACACCTAAA  840
281  ProSerLeuIleLysAlaLeuAlaLeuIleAlaThrAaspValGlyLeuGlyTyrPro  300
841  CTTCTCTTAATAAAAGCTGCACCTTATCGCTGGGCTACTGATGTGGTTTAGGATATCCA  900
301  SerGlyAaspGlnGlyTrpGlyArgValThrLeuAaspLysSerLeuAenValAlaTyrVal  320
901  AGTGTGACCAAGCTGGGGCGGTGTTACTCTAGTAAATCGTTAAATGTAGCGTATGTC  960
321  AenGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla  340
961  AATGAAGCAACTGCATTTAGCAACAGCAAAAGCAACGTATTCGTTCCAAAGCAAGCG  1020
341  GlyLysProLeuLysIleSerLeuValTrpThrAaspAlaProGlySerThrThrAlaSer  360
1021  GGTAAACCTTTAAAAATCTCGTTAGTATGGACAGATGTCACCTGGAGTACAACTGCACT  1080
361  TyrThrLeuValAenAaspLeuAaspLeuValIleThrAlaProAenGlyGlnLysTyrVal  380
1081  TATACACTAGTTTAATGATTTAGATCTAGTTATTACTGCTCCGAATGGACAAAATATGTA  1140
381  GlyAenAaspPheSerTyrProTyrAaspAenAenTPAaspGlyArgAenAenValGluAen  400
1141  GGAAATGATTTTAGTTATCTTATGATTAATCACTGGGATGGTCCGCAACAATGTTGAGAAC  1200
401  ValPheIleAenAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAenVal  420
1201  GTATTATTAAACGCTCCGCAATCTCGAAGCTATACAAATTGAGTTCAAGCGTATAATGTA  1260
421  ProSerGlyProGlnArgPheSerLeuAlaIleValHis  433

```

```

Db      1261 CCAATGTCGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299
RESULT 2
AB046402
LOCUS      AB046402      1299 bp      DNA      linear      BCT 16-JUN-2001
DEFINITION Bacillus sp. D6 PROA gene for protease, partial cds.
ACCESSION AB046402
VERSION   AB046402.1 GI:12381936
KEYWORDS
SOURCE    Bacillus sp. D6
ORGANISM  Bacillus sp. D6
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (sites)
AUTHORS   Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
           Horikoshi,K.
TITLE     Novel oxidatively stable subtilisin-like serine proteases from
           alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
           evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
JOURNAL
MEDLINE   20568675
PUBMED    11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS   Saeki,K.
DIRECT SUBMISSION
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2605, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
Location/Qualifiers
1..1299
/organism="Bacillus sp. D6"
/mol_type="genomic DNA"
/strain="D6"
/db_xref="taxon:127889"
1..1299
/gene="PROA"
<1..>1299
/gene="PROA"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21265.1"
/db_xref="GI:12381937"
/translation="NDVARGIVKADVQNNQGLYGGQGVAVADTGLDTRNDSSMHE
AFRGKITALYALGRNTNNDPNHGTHVAGSVLGNLKNKMAPOANLVFQSIMDSSGG
LGGLPSNLTLPQAWNAGARLHTNSWGA PVNNAVYANSROVDVVRNNDMTVLPAG
NEGNSGISAPGTAKNALTYGATENYSPSGSIADNPNHIAQSSRGATRDGKPD
VTAPGTFILSARSSLAPSSFWANTNSKYAYGSGTSMATPIVAGNVAQLRHFYKRG
ITPKPSLIKAALDIAGATDVGLGPGSDQDQGRVTLDKSLNVAIVNEATALTIGQATY
SFQTAGKPLKISLVMTDAPGSTTASYTLVNDLDLIVITAPNGQKYGVNDFSYPYDNNW
DGRNNVENVFINAPQSGVTIEVQAYNPVSPGQRFSLAIV"
ORIGIN
Alignment Scores:
Pred. No.: 2,34e-133 Length: 1299
Score: 2251.00 Matches: 430
Percent Similarity: 99.54% Conservative: 1
Best Local Similarity: 99.31% Mismatches: 2
Query Match: 99.56% Indels: 0
DB: 1 Gaps: 0
US-09-985-689A-5 (1-433) x AB046402 (1-1299)
QY      1 AsnAspValalaArgGlyIleVallysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db      1 AATGATGTAGCAGAGGAATAGTAAAGAGCAGAGCTTGACAAAACAAATTACGGACTATAT 60
QY      21 GlyGlnGlyClnValValAlaValAlaAspThrGlyLeuAspThrGlyValArgAsnAspSer 40
Db      61 GGCAAGGTCAGTAGTTCGATGCGGATACCGGTTTATAGATACAGGTCTGATACGATAGT 120
QY      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

```

```
Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 CCAATCGGACCAGCGTTCTCACTAGTATCGTACAT 1299

RESULT 3
E03808
LOCUS      2218 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION      DNA encoding alkaline protease Ya.
ACCESSION      E03808
VERSION      E03808.1 GI:2172022
KEYWORDS      JP 1992197182-A/1.
SOURCE      Bacillus sp.
ORGANISM      Bacillus sp.
REFERENCE      1 (bases 1 to 2218)
AUTHORS      Tobe, S., Odera, M. and Asai, Y.
TITLE      DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
JOURNAL      PROTEASE YA USING THE DNA
PATENT      JP 1992197182-A 1 16-JUL-1992;
LION CORP
COMMENT      OS      Bacillus sp.
PN      JP 1992197182-A/1
PD      16-JUL-1992
PE      28-NOV-1990 JP 1990327110
PI      TOSE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
PC      C12N15/57, C11D3/386, C12N9/54, (C12N15/57, C12R1/07), (C12N9/54,
PC      C12R1/07);
CC      strandedness: Double;
CC      topology: Linear;
CC      hypothetical: No;
CC      anti-sense: No;
CC      *source: strain=Y strain;
CC      *source: clone=pUB8A;
FH      Key
PH      Location/Qualifiers
FT      CDS      218..2125
FT      /product='precursor of alkaline protease Ya'
FT      sig_peptide      218..823
FT      mat_peptide      824..2122
FT      /product='Precursor of alkaline protease Ya'.

FEATURES
source      Location/Qualifiers
1..2218      /organism='Bacillus sp.'
/mol_type='genomic DNA'
/db_xref='taxon:1409'

ORIGIN
Alignment Scores:
Pred. No.:      1,06e-132      Length:      2218
Score:      2245.00      Matches:      429
Percent Similarity:      99.77%      Conservatives:      3
Best Local Similarity:      99.08%      Mismatches:      1
Query Match:      99.29%      Indels:      0
DB:      6      Gaps:      0

US-09-985-689A-5 (1-433) x E03808 (1-2218)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnTyrGlyLeuTyr 20
Db 824 AATGATGTAGCAGAGGATAGTAAAGCTGATGTCACAAAACAAATACCGATATAT 883
Qy 21 GlyGlnGlyGlnValAlaAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGACAAGGTCAACTAGTTGTCAGTAGCGGACACAGGCTTAGATACAGTCTACAGTAGT 943
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 944 TCTATGATGAGGATTCGCGGGGAATACAGCTCTTTACCGGTTAGGAAGAACTAAT 1003
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1004 AATGCGAGTATCGAATGGCATGGCATGGCAGACACATGTAGCAGGTTCTGTCTTGGTAATGCT 1063

Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1064 TTTAATAAAGGAATGGCTCGCAGCTACTTAGTCTTCCAATCTATTATGGATAGCAGC 1123
Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 1124 GAGAGGATTAGTGCTTACCATCGAACTTAAATACGTTATTTAGTCAAGCTCGAATGCT 1183
Qy 121 GlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 1184 GAGACAGAATTCATATACTCTTGGGAGCCCCAGTAAATGGAGCGTACACTGCTAAC 1243
Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1244 TCGAGACAGTGGATGATGATGTTCCGAAATAATGATATGACGCTACTTTTTCAGCTGCT 1303
Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1304 AATGAGGTCTTAATTCAGGAACAATAGTCTCCAGGTACAGCGAAATAATGCTATTACG 1363
Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
Db 1364 GTCCGCGCAACCGAAACTATCGCCAAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 1423
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 1424 ATTGCACAATTTTCATCGAGAGGACTACGAGGATGGAGCAATTAAGCCTGACGTAAACA 1483
Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTTP 240
Db 1484 GCTCTCGAAACATTTATTTATCAGCAGCTTCTTCTTAGCTCCAGACTCTTCGTTTGG 1543
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 1544 GCGAATTAACAGTAAATACGCTATATGGCGGACCTCCCTCCATGGCAGCACCTATGTT 1603
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 1604 GCAGGGAATGCGCGCAATTCAGTGAGCAATTTTATAAAAAATAGAGGTATTACTCTTAAG 1663
Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 1664 CCTTCCTTTAATAAAGCTGCACCTTATCGCTGGTGCTACTGATGTGGTTTAGGATATCCT 1723
Qy 301 SerGlyAspGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
Db 1724 AGTGGTGACCAAGCTCGGGGCGGTACTCTAGATAAATCGTTAAATGTAGCGTATGTC 1783
Qy 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
Db 1784 AATGAAGCAACTGCATTAGCCACAGGACAAAAGCAACGATTTCGTTCCAGCACAGCG 1843
Qy 341 GlyLysProLeuLysIleSerLeuValTTPThrAspAlaProGlySerThrThrAlaSer 360
Db 1844 GGTAAACCTTTTAAATAATCTCGTTAGTAGACAGATGCTCTCTGGAAGTACACTGCATCT 1903
Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 1904 TATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGAAATGGACAAAATATGTA 1963
Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
Db 1964 GGAATGATTTTAGTTATTCCTTATGATTAATAACTGGGATGGTGCACAACTATTGGAGAC 2023
Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 2024 GTATTTATAACCGCTCCGCAATCTGGAACGTATATAATTCAGGTTCAAGCGTATAATGTA 2083
Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 2084 CCATCTGCCACAGCGTTTCTCACTAGCTATCGTACAT 2122
```

```

RESULT 4
AB046404
LOCUS       Bacillus sp. Y PROC gene for protease, partial cds.
DEFINITION
ACCESSION  AB046404
VERSION     AB046404.1
KEYWORDS   G1:12381940
SOURCE     Bacillus sp. Y
ORGANISM   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE  1 (sites)
AUTHORS    Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
            Horikoshi,K.
TITLE      Novel oxidatively stable subtilisin-like serine proteases from
            alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
            evolutionary relationships
JOURNAL    Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE    20568675
PUBMED     11118284
REFERENCE  2 (bases 1 to 1299)
AUTHORS    Saeki,K.
DIRECT SUBMISSION
SUBMITTED  (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
            Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
            Tochigi 321-3497, Japan (E-mail:387185@kcastanet.kao.co.jp,
            Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES   Location/Qualifiers
            1..1299
             /organism="Bacillus sp. Y"
             /mol_type="genomic DNA"
             /strain="Y"
             /db_xref="taxon:133779"
            1..1299
             /gene="PROC"
             <1..>1299
             /gene="PROC"
             /evidence=experimental
             /transl_table=1
             /product="protease"
             /protein_id="BAB21267.1"
             /db_xref="GI:12381941"
             /translation="NDVARGIVKDAVQNTGLVGGQGVVAVADTGLDTRNDSSMHE
             AFRGKITLALYLRNTASDPNGHGVAGSVLGNALNKGVAPOANLIVFQSDSSGG
             LGLEPSNLNTLFSQAWNAGARIHTNSWGVNAGVYANSROVDYVRNNDMTVLFAAG
             NEGPSNGTISAPGTAKNAITVGATENYRSPFGSADNPNIHQFSSRGATDRGIRKPD
             VTAPTFLSARSSLAPOSSFWANYSVAYMGVTSMTPIVAGNVQLRHFVKNG
             ITPKSLIKALIAAGATDVLGYPNGDQMGVTLNKSIVAYVNEATALATGQKATY
             SFQAQAGPLKISLWTDAPGSTTASYTIVNDLIDLIVTAPNGQKIVGNDFFSYPDNNW
             DGRNNVENFINAPSGTYIIIEVQAINVPSGPQRPFLAIVH"
ORIGIN
Alignment Scores:
Pred. No.: 1,16e-132 Length: 1299
Score: 2240.00 Matches: 428
Percent Similarity: 99.77% Conservative: 4
Best Local Similarity: 98.85% Mismatches: 1
Query Match: 99.07% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-5 (1-433) x AB046404 (1-1299)

Qy 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGGTAGCAAGAGGATAGTAAAGCTGATGTTGCACAAAACAATTACGGATTATAT 60
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GCACAGGTCAAGTAGTGTGACATAGCGGACACAGCTTAGATACAGGTCTGTAACGATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGATGAAGACATTCGCGGGGAAATCACAGCTCTTTACGCGCTTAGGAAGAACTAAT 180

61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
181 AATGGAGTATCGAATGGCATGGCAGATGACACATGTAGCAGGTCTCTGACTTGTATGCT 240
Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 241 TTAATAAAGGAATGGCTCGCAAGTAATCTTAGCTTCCAAATCTATTATGGATAGCAGC 300
Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
Db 301 GGAGGATTAGTGGCTTACCATCGAATTAATACGTTATTTAGTCAAGCTTGGATGCT 360
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 361 GGAGCAAGAAATTCATACCTAATCTTGGGAGCCCAATGAGTAAATGAGGAGTACACTGCTAAC 420
Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 421 TCAGACAGTGGATGATGATATGTTCAATATATGATATGACGTACTTTTGCAGCTGGT 480
Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 481 AATGAAGTCTTAATTCAGGAACAATAGTGTCTCCAGGTACACGCAAAAATGCTATTACG 540
Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
Db 541 GTCGCGCAACGGAACACTATCGCCAGCTTCGGTTCGATAGCAGTACACCCAAATCAT 600
Qy 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 601 ATTGCACAATTTTCATCGAGGAGCTACGAGGATGGACGAATTAAGCTGACGTAAACA 660
Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 661 GCTCCTGGAACTTATTTATCAGCACGTTCTTCTTAGCTCCAGACTCTTCGTTTGG 720
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 721 GCGAATTAATAACAGTAATAACGCGTATATGGCGGTACCTCCATGCGCACACCTATTGTT 780
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 781 GCAGGGAATTCGCGCAATTTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCTAAG 840
Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 841 CTTCTTTAATAAAGCTGCACCTTATCGCTGGTGTCTACTGATGTTGGTTAGGATATCCT 900
Qy 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
Db 901 AATGGTGCACAGGCTGGGGCGTGTACTCTAAATAAATCGTTAAATAGCGTATGTC 960
Qy 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
Db 961 AATGAAGCAACTGCATTCAGCCAGACAGCAAAAAGCAACGTTTTCGTTCCAGCACAGCG 1020
Qy 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
Db 1021 GGTAAACCTTTAAATAATCTCGTTAGTATGCACAGATGCTCCTGGAAGTACAACTGCATCT 1080
Qy 361 TyrThrIleuValAsnAspLeuAlaIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 1081 TATACACTAGTTAATGATTAGTATCTAGTTATCTGCTCCGAATGGACAAAATATGTA 1140
Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
Db 1141 GGAATGATTTTAGTTATCTTATGATATAATACTGGATGGTGGCAACAATGTTGAGAAC 1200
Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 1201 GTATTATTAACGCTCCGAATCTGGAACGATATAATTAATTTGAGGTTCAAGCGTATATGTA 1260

```



```
Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
|||||
Db 1261 CCATCTGGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 5
AR069954
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1. (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
LOCATION/Qualifiers
1..3003
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 7,24e-120 Length: 3003
Score: 2044.00 Matches: 386
Percent Similarity: 94.69% Conservatives: 24
Best Local Similarity: 89.15% Mismatches: 23
Query Watch: 90.40% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-5 (1-433) x AR069954 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValAlaAspValAlaGlnAsnSerGlyLeuTyr 20
|||||
Db 1470 AATGACGTGCCCGTGGCATTTGTGAAGCAGACGTGCGACAAATAACTTTGGCTATAT 1529

Qy 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
|||||
Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATCTGGCTTATACAGGAAGATGACGT 1589

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
|||||
Db 1590 TCGATGATGAAGCATTCGCGGTGAAGATTACGCGCATATATGCTGGCGAGCAAGAA 1649

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
|||||
Db 1650 AACCCCAATGATCCAAATGGACATGGAAACCAATGTTGCTGATCTGTTAGGAATGCT 1709

Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
|||||
Db 1710 ACAATAAAGGATGGCAGCCAGCAAGCAATCTAGTCTTCAATCTATTATGGATAGTGT 1769

Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTTPAsnAla 120
|||||
Db 1770 GGAGGCTGGGAGACACCTCTGCTAATCTACAAATATTTCAGTCAAGCATATAGTCT 1829

Qy 121 GlyAlaArgIleHisThrAsnSerTTPGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
|||||
Db 1830 GGAGCGAGAAATTCATACGAATTCATGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889

Qy 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
|||||
Db 1890 TCTCGAATGTGTATGATATTGTGAGAAAAAATGATATGACGATCTCTTTTGGCGCGGA 1949

Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
|||||
Db 1950 AATGAGGACACGAGTGGGTACAAATCAGTGCACCAAGCAAGCAAAATGCGATTACA 2009

Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
|||||
Db 2010 GTTGGGCGACCGAAACCTACGTCCCAAGCTTCGGATCTTATCGGATAATATTAACCAT 2069
```

```
Qy 201 IleAlaGlnPheSerSerArgGlyValAlaThrArgAspGlyArgIleLysProAspValThr 220
|||||
Db 2070 GTTGTCTCAATTCCTTCAAGAGGTCTTACTAGAGATGAGCGTATTAAAGCCGAGCTCATG 2129

Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
|||||
Db 2130 GCACGAGTACGTATATTCCTCTGCTAGATCATCATTTAGTCCAGATTCCTCATTTCTGG 2189

Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
|||||
Db 2190 GCAACCATGATAGTAAATATGCTACATGGTGGTACTTCTATGGCTACTCCAAATGTA 2249

Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
|||||
Db 2250 GCAGGTAAATGTGCACAAATTAAGGAGCATTTTGTGAANAATAGAGGGGTAACTCTTAAG 2309

Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
|||||
Db 2310 CTTCCCTTTTAAAGGCTGCTTTAATTGCAGGTGCTCGGATGTTGGACTTGGCTTTCCA 2369

Qy 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
|||||
Db 2370 AATGTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTGCAATTTGTG 2429

Qy 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
|||||
Db 2430 AATGAACGAGCCTTTATCAACAGTCMAAAGCAACATATTCGTTTACGGCTCAAGCT 2489

Qy 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360
|||||
Db 2490 GGTAACCCCTTAAATAATATCACTTGTGTTGTCAGATGCCAGGTAGCAGCGCATCA 2549

Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
|||||
Db 2550 CTAATTTAGTGAATGATTAGACTTTAGTATCTAGTCCCAATAGAACTAAATATGCTC 2609

Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsn 400
|||||
Db 2610 GGAAATGACTTACAGCACCGCTATGATAACAATTTGGATGGCAGAAACAACGCGGAAAT 2659

Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
|||||
Db 2670 GTGTTTATCAATGCTCTCAAGCGGAACGATATACGTCGAAGTGCAGGCTTACAATGTA 2729

Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
|||||
Db 2730 CCAGTAAGTCCGCAACCTTTTCTTTAGCGATTGTACAT 2768

RESULT 6
BD062155 3003 bp DNA linear PAT 27-AUG-2002
LOCUS BD062155
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:23607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
REFERENCE
1. (bases 1 to 3003)
AUTHORS Sloma, A. and Christianson, L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: JP 2001514529-A 39 11-SEP-2001;
COMMENT
NOVO NORDISK BIOTECH INC
PN JP 2001514529-A/39
PD 11-SEP-2001
PF 09-JUN-1998 JP 1999503145
PI 12-JUN-1997 US 08/873479
PC C12N15/57, C12N15/75, C12N9/54, C12K14/00
CC Strandedness: Single;
Topology: Linear;
```

FEATURES

QY FH Key Location/Qualifiers.
source 1..3003
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores: 7.24e-120 Length: 3003
Pred. No.: 2044.00 Matches: 386
Score: 94.69% Conservative: 24
Percent Similarity: 89.15% Mismatches: 23
Best Local Similarity: 90.40% Indels: 0
Query Match: 6 Gaps: 0
DB:

US-09-985-689A-5 (1-433) x BD062155 (1-3003)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGCGGCGATTTGAAAGCAGACGTCGCAAAATAACTTTGGCTTATAT 1529
QY 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGCACAGGACAGATTGTAGCAGTTGCTGACTGGCTTGATACAGAGAAATGACACT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGCATGAAGCATTCGCGGTAAAGATTACCGCATATATGCTGGCGCAGAAAGCAAT 1649
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1650 AAGCCCAATGATCAAAATGACATGGAAACCAATGTTGCTGGATCTGTGTAGGAATGCT 1709
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1710 ACAATTAAGGATGGCAGCGCAAGCAATCTAGTCTTTCAATCTATTATGATAGTGGT 1769
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
Db 1770 GGAAGGCTGGGAGGACTACTGCTAATCTCAACAACATTATTCAGTCAAGCATATAGTCT 1829
QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyValAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 1830 GGAGCGAGATTTCATACGAATTCATGGGGGCTCCAGTAAACGGTCCCTATACGACAGAC 1889
QY 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1890 TCTCGAAATGTTGATGATTATGTGAGAAATAATGATAGCAGATCTTTTTCGGCGCGGA 1949
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1950 AATGAGGACCGAGTAGCGGTACAAATCAGTGCACCGAGAACAGCAAAATGCGATTACA 2009
QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
Db 2010 GTTGGGCAACCGAAACCTACGTCCAAAGCTTCGGATCTTATCGGATATATATACCACT 2069
QY 201 IleAlaGlnPheSerSerArgGlyValAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 2070 GTTCTCAATCTCTTCAGAGGTCCTAGATGAGATGAGATTAAGCCGCGAGCTCATG 2129
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
Db 2130 GCACCGGTACGTATATCTCTCTGTAGATCATCATAGTCCAGATTCCTCATCTCTGG 2189
QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 2190 GCAACCATGATAGTAATATGCGCATGCGGTGCTACTCTTATGGCTACTCCCAATTGTA 2249
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 2250 GCAGGTAATGTTGCACAAATTAAAGGAGCATTTTGTGAAAAAATAGAGGGGTAACTCTCTAAG 2309

RESULT 7

AB046406 1302 bp DNA linear BCT 23-JAN-2001
LOCUS
DEFINITION Bacillus sp. NV1 PROBE gene for protease, partial cds.
ACCESSION AB046406
VERSION AB046406.1 GI:12381944

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

gene

CDS

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

Location/Qualifiers

1..1302

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic Bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

11118284

2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUN-2000) Katsuhisa Saeki, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kastonet.kao.co.jp,

/db_xref="GI:12381945"
/translation="NDVARGIVKADVAQSSVGLYGGGVAVADTGLDTRNDSSMHE
AFRKLITAIYLFQSRAGRIHNSGAPVNGAITIDSRVDYIVKNDVAVLFAA
GLGLPSNVSTLFSQIAKNAITVSGVADNINHVAFSSGPTGDKRIK
GNEGNGTISAPGTAKNAITVSGVADNINHVAFSSGPTGDKRIK
DYMAPGTIFLSAPSSVAFWANDSKYMGSTWATPIVAGNVAQLRHEFIK
GITPKSLKALIAIGATIGLGYPSGNGWRLDKSLNVAFTNETSSLTQKAT
YFTAGSGKPLKISLVSDAPASTSASVTLNDLVLITAPNGTKYVGNDFAPYDNN
WDGRNNVNFVINAPOSGTYVEVQAYNVDPQPAFLAIVN"

ORIGIN

Alignment Scores: 4.57e-119 Length: 1302
Pred. No.: 2024.50 Matches: 384
Score: 94.93% Conservatives: 28
Percent Similarity: 88.48% Mismatches: 21
Best Local Similarity: 89.54% Indels: 1
Query Match: 1 Gaps: 1
DB:

US-09-985-689A-5 (1-433) x AB046406 (1-1302)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AACGATGTTGCCAGAGCAATGTAAAGCCGATGTTGCCAGAGCACTATGTTTATAT 60
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGGCAAGGGCAAGTGGTTCAGTAGCGGATACCGGACTGGATACAGGGCGTAATGACAGC 120
QY 41 SerMetHisGluAlaPheAeGlyLysIleThrAlaLeuValAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAGCGTTCGAGGAGAAAGATTACCGGATATATGCCCTTGGAGACAAAC 180
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 181 AACGCCAATGATCCAAACGGACACCGGACGCGATGTTCCGGATCTGTTTAGGAACGGT 240
QY 81 Leu---AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 ACTTCGAATAAGGATGGCTCCCAAGCTAAGTATGTTTCCATCTGTATGTCGACG 300
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 301 AATGGCGGACTGGCGGACTGCCITCCAAATGTAAGTACATATTATTCAGCCAGGCATATAGT 360
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAlaGlnGlyValAlaTyrThrAla 139
Db 361 GCCGGTGCAGAAATCCATACGAATCATGGGAGCGCTGTAACGGAGCGGTACACTACT 420
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAACGCTAGACGATATGTTCTAAATAATGATATGGCGGTCTCTTTTGCAGCG 480
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 CGTAAACGAGGCGCAATGGCGGCAATCATGTCCTCTGTCGACCGGAGAAATGCTATC 540
QY 180 ThrValGlyValAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 541 ACAGTAGGGGCACAGAAACCTCGCCCAAGCTTTGATCTTATCTGTCGACACATCAAT 600
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTAGCACAGTCTTCTCCCGGACCTCAAAAGGATGGACGATCAAAACCGGACGTA 660
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCACACGAGCAATTTATTTATTCGCAAGATCTTCTTGGCTCCGAGCTCTCTTCATTC 720
QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 721 TGGGCAACCATGACACGCAATATGTTATATGCGTGGAAACATCATCCATCGGACACCGGAT 780

QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTAGCTGGTAACCTTGACAGTTACGTGAACATTTTCATCAAAACAGAGGAATCACTCTC 840
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyValAlaThrAspValGlyLeuGlyTyr 299
Db 841 AACCATCTCTTGTGAAGAGAGCTCTTATTGCGGAGCACTGATATCGTCTTGGCTAT 900
QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 901 CCGAGTGGAAACCAAGGATGGGAAGAGTAGTAACTTGGACAAGTCATTAATGATGCTTTC 960
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 961 GTAAATGAACACAGCTCTTTTATCTACTACCAAAAGGCTACGTTATTCATTACTGCACAA 1020
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1021 TCAGGCAAAACCTTTGAAGATTTTCATGTTGGTCTGATGACCGGCAAGTACTTCGCA 1080
QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCGGTTACATTTGGTGAATGATCTGGATCTGGTGAATACAGTCCAAATGGACAAAGTAT 1140
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTTGGAAACGACTTTACTGCTCCCTATGATAATACTGGGATGACGCTAACATGTAGAG 1200
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1201 AACGTGTTTATCAATGCTCCGCAAGCGGACGATATACAGTTGAGGTACAGGCTTACAT 1260
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTTCCACAGGGCCGCGCGGCTTTCTTTGGCTATTGTGAAC 1302

RESULT 8

AB084155 1923 bp DNA linear BCT 02-SEP-2003
LOCUS Bacillus sp. KSM-9865 gene for protease, complete cds.
DEFINITION
AB084155
ACCESSION
AB084155.1 GI:34392386
VERSION
AB084155.1
KEYWORDS
SOURCE
ORGANISM

REFERENCE
1
Okuda, M., Saeki, K. and Kobayashi, T.
Bacillus sp. KSM-9865 protease gene
Published Only in Database (2003)
2 (bases 1 to 1923)
Okuda, M., Saeki, K. and Kobayashi, T.
Direct Submission
Submitted (18-APR-2002) Mitsuyoshi Okuda, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail: okuda.mitsuyoshi@kao.co.jp,
Tel: 81-285-68-7543, Fax: 81-285-68-7547)

FEATURES
source
1..1923
/organism="Bacillus sp. KSM-9865"
/mol_type="Genomic DNA"
/strain="KSM-9865"
/db_xref="taxon:192495"
1..1923
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAC82522.1"
/db_xref="GI:34392387"

CDS

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

1..1923

AOSSGYLYGOGQIVAVADTCLDTGRNDSMHEAFRCIKITALLYALVAGCTNNANDNGHGT
 HVAAGVNGSTNGKNAPOANLVFQSIMDSGGGLGGLPSNLQTLTSQASAGARHTN
 SWGAANVAGTIDNRKNDMTILFAENGFGNGTISAPGTAKNALIVGATE
 NLRPSFGSADNHNHVAQFSSRGTDKRIPDVMAPGTFILSARSLSPDSSFWANH
 DSKYAYMGTSMTPIVAGNVAQREHFVKNRGRTIKPSLLKALITAGAADIGLGYPN
 GNGWRVITDKSLNWAYNNESSLSSTSQKATYSTFTATGPKLISLWSDAPASTTA
 SVTLVNDLDLVIAPNGTQVVGNDFTSPYNNWDRNNVENFVFNAPQSGTIVIEQA
 YNVPVGPQTSALVN"

ORIGIN

Alignment Scores:
 Pred. No.: 1,53e-117 Length: 1923
 Score: 2003.50 Matches: 382
 Percent Similarity: 94.24% Conservative: 27
 Best Local Similarity: 88.02% Mismatches: 24
 Query Match: 88.61% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689A-5 (1-433) x AB084155 (1-1923)

QY 1 AsnAspValAlaAAGGlyVleValLysAlaAspValAlaAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 619 AATGATGTTGACGTGGAATGTCAAGCGGATGTGGCGCAGACGACGTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAGGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGGTCCCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyVleThrAlaLeuTyrAlaLeuGlyVArgThrAsn 60
 Db 739 TCGATGATGATGAAGCTTCCGGGGGAAATTAATCTGATATATGATTCGATTCGGACGACGAT 798
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATGCCAATGATACGAATGGTCATGCTACGATGCTGGCTCGGTATAGGAAACGGC 858
 QY 80 AlaLeuAsnTyrGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTATCATGGAATAGC 918
 QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
 Db 919 GGTGGGGACCTTGAGGACTACCTCGAATCTGCAACCTTATTCAGCAAGCATACAGT 978
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValTyrAla 139
 Db 979 GCTGTGCCAATTCATACAACTCTGGGAGCAGCGATGGAATGGGCTTACACACA 1038
 QY 140 AsnSerArgGlnValAspGlyLysValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACATCTCTTTTCGTGCC 1098
 QY 160 GlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGATGAGGACCGAAGCGGGAACCATCATGTCACCAAGGATGACGATCTTAAATA 1158
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACATGCGAGNCTACGGAAACCTCGCCAGCTTCGGGTCTTATGCGGACAAATATCAAC 1218
 QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyVArgIleLysProAspVal 219
 Db 1219 CATGTGGCACAGTCTCTTCATCGTGGACCGACAAAGGATGGACGATCAAAACCGGATGC 1278
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
 Db 1279 ATGGACCGGGAACTTCATATATCATCAGCAAGATCTTCTGTGACCGGATTCCTCTTC 1338
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGGCGAACCATGACAGTAATATGATCATATGATGGGTGGAACTCCATGCTACACCGATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279

Db 1399 GTTGTGGAACGTGGCACACCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
 QY 280 LysProSerLeuIleLysAlaAlaIleuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCCTTCTTATTAAAGCGGCACTGATTCGCCGTGCGAGACATCGGCTTGGCTAC 1518
 QY 300 ProSerGlyAspGlnGlyTyrGlyVargValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATGGATAAATCCCTAAACGTTGCCTAT 1578
 QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGAGTCCAGTCTCTATCCACCACCAAGCGAGCTACTCGTTTACTGCTACT 1638
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrAla 359
 Db 1639 GCCGCAAGCCTTTGAAGATCTCCCTGATGCTGATGCTGCGGACCAACTGCT 1698
 QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1699 TCCGTAACGCTTGTCAATGACCTGGACCTTGTCAATACCGCTCCAAATGGCACCAATAT 1758
 QY 380 ValGlyAsnAspPheSerTyr-ProTyrAspAsnAsnTyrAspGlyVargAsnValGlu 399
 Db 1759 GTTGAAGATGACTTTACTTCGCCATACATAATACTGGGATGGCGCAATAACGTAGAA 1818
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGTATTATTATATCGCCACAAAGCGGAGCTATACAAATGAGGTACAGCTTATAAC 1878
 QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGTTGGACCAACACCTTCTCGTTGGCAATGTGAAC 1920
 RESULT 9
 LOCUS AB046403 1920 bp DNA linear BCT 10-MAY-2002
 DEFINITION Bacillus sp. 9860 PROA gene for protease, complete cds.
 ACCESSION AB046403
 VERSION AB046403.2 GI:20521152
 KEYWORDS Bacillus sp. 9860
 SOURCE Bacillus sp. 9860
 ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1
 AUTHORS Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 11118284
 REFERENCE 2 (bases 1 to 1920)
 AUTHORS Saeki, K.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail: 387185@kasetanet.kao.co.jp, Tel: 81-285-68-7400, Fax: 81-285-68-7403)
 COMMENT On May 9, 2002 this sequence version replaced gi:12381938.
 FEATURES
 source
 1. .1920
 /organism="Bacillus sp. 9860"
 /mol_type="genomic DNA"
 /strain="9860"
 /db_xref="taxon:133778"
 1. .1920
 /gene="PROA"
 1. .1920
 /gene="PROA"
 /codon_start=1
 gene
 CDS

```

/transl_table=11
/product="protease"
/protein_id="BAB21266.2"
/db_xref="GI:2521153"
/translation="WRKKKVFSLVLSAAIILSTVALNPPSAGDARTFDLPKGIQTIT
DVSFKQRGTAAAFVSESNKLLKGLKLETVPANNNKLHIVQNGPFLIEIKOK
LETTGAKLIDYIPDYATVEVEGDVQSVRIEIVSVESVEPYLPKIKDIPQFTKGAST
LVKALALDITQNNKEVQVQRIEIEIAQYVSNDRVHYITAKPEYKVMNDVIRKADVA
QSSVLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRITNNANDTNGHGH
VAGSYLGNQKNGKMAPOANLVFQSIMDSGGGLGPNLOTLFQAQPSAGARITHNS
WGAANGVATYDTSRNVDDYVRKNDMTILPAAGNERPNGGTISAPGTAKNAITVGCATEN
LRPFGSADNINVAQSSRRGPTQDRIKIDPMAPGYIILSARSSAPQSSFWANDH
SKYAYNGGSMATPIVAGNVAQLRHFVKNRGITIPKPSLLKHALIAGNADVGLGYPNG
NQWGRVLDLKLNAVYVNESALSQKATVTFATGAKPKIKSLVMSDDAPASITTS
VTLVNDLDIVITPAGNTRYVGNDFAPFDNNWNGRNVNFIINSPOSGTYTIEVQAY
NVPVGPQNPFLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 1,77e-117 Length: 1920
Score: 2002.50 Matches: 383
Percent Similarity: 94.24% Conservative: 26
Best Local Similarity: 88.25% Mismatches: 24
Query Match: 88.57% Indels: 1
DB: 1
Gaps: 1
US-09-985-689A-5 (1-433) x AB046403 (1-1920)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGGTAATTGTCAAAAGCGGATGTGGCACAGACGACTACGGTTTGTAT 675
QY 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGCGCAGATGTCGAGTGTCCGATCTGGATTGGATGGATACAGAGAAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGATGAAGCCTCCGCGGTAAATAACAGCACTATATGCACTGGGTCCGAGMAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCGAATGATACGAACGGTCTGTTACCATGTCAGGTTCGGATTATAGGAATGTC 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGATTAAGGATGGACCTCAAGCGAATCTGGTTTTCATCCATCATGTGATACG 915
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 916 AGTGGTGGGCTTGGAGGCTTGCCCTTCCAACTCGCAACACCTATTTCAGCCAAAGCAATCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaProValAsnGlyAlaTyrThrAla 139
Db 976 GCAGTGCCAGAAATTCATACAACTCTGGGGGCGACGGTGAATGGGCTTACACGACA 1035
QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1036 GATTCCAGAAATGTGGATGATCTATGTAAAGGAAATGATATGACGATCTTTTCGGCGGT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGAAATGAAGCCGACCGCGGTACCATGATGTCACCTGGTACGGCTAAAACGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1156 ACAGTGGCGCAACCGAAACCTGGCTCAAGCTTCGGTTCTCTATGCGAGATAATATTAC 1215
QY 200 HistLeuAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1216 CACGTTGCACAGTCTCTCTCCGCGGCCCGCAAAAGATGGCGGAATCAAGCCCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239

```

1276 ATGGCCCGCCAGGACATACATATTTTATCAGCAAGATCTTCTCTTGCCCGCATTCCTCTTC 1335
240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro 259
1336 TGGCGCAATCATGACAGCAAAATATGCTATATGGGTGGAAACGTCCTCATGCGCAACCGCAT 1395
260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
1396 GTTTCGGGGGAATTTGCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGAACTACTCTCT 1455
280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
1456 AAGCTTCCCTATTGAAGACAGCTTTGATTGAGGTGCTGCTGATGTTGGATTGGGTTAT 1515
300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
1516 CCGAACGGAACCAAGGATGGGCGCGAGTGACCTCGATGAATTCGTTGAACGTTGCTTAT 1575
320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
1576 GTGACGATCCAGTGCCTATCACTAGCAAAAGCGACATATACCTTTACTGCAACG 1635
340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrAla 359
1636 GCGGCGAAGCCATTGAAATCTCCCTGGTATGTCGATGCGCCCTCGAAGCATCTACTGCT 1695
360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
1696 TCTGTAACTTGGTCATGATTTGGATTGGTATTACAGCAACCAACGCAAGATAT 1755
380 ValGlyAsnAspPheSerTyrProTyrAspAsnThrAspGlyArgAsnValGlu 399
1756 GTCGGAATGACTTCTCAGCACCATTTTGACAAATACCTGGGATGCGCGCAATAACGTAGAA 1815
400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
1816 AATGATTTATTAATTCGCCCAAGTGGACATATACCATTCAGGTGCGCAAGCATATAAT 1875
420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
1876 GTGCGGTTGGACCAAAACTTCTCGTTGGCAATTGTGAAC 1917

RESULT 10
AR368116 1920 bp DNA linear PAT 12-SEP-2003
LOCUS Sequence 3 from patent US 6376227.
DEFINITION AR368116
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
Shikata,S. and Nomura,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1,77e-117 Length: 1920
Score: 2002.50 Matches: 383
Percent Similarity: 94.24% Conservative: 26
Best Local Similarity: 88.25% Mismatches: 24
Query Match: 88.57% Indels: 1
DB: 1

US-09-985-689A-5 (1-433) x AR368116 (1-1920)

Db 979 GCTGTGCGCAGAAATTCATACAACTCTCTGGGACGACGAGTGAATGGGCTTACACA 1038
 Qy 140 AnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGGATGACTATGTGCGCAAAATGATGACGATCTTTTCGGTGCC 1098
 Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleVal 179
 Db 1099 GGGAAATGAAGACCGAAGCGGGAACCATCATGTCACGAGCAGCATAAATGCAATA 1158
 Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnPro 199
 Db 1159 ACAGTCGAGAGCTACGGAACACTCGCCCAAGCTTTGGGCTTATGGGCAATATCAAC 1218
 Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleTyrProAspVal 219
 Db 1219 CATGTGGCAGCACTCTCTTCACGTGCGGACCGCAAGGATGACGATCAAAACGGATGC 1278
 Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
 Db 1279 ATGGCAGCGGAGAGCTTATCATCTATCAGCAAGATCTTCTCTGACCGGATCTCTCTTC 1338
 Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGGCGAACCATGACAGTAATATGCTATACATGCTGGTGAACCTCCATGGCTACACGATC 1398
 Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTGTGGAACGCTGCGACACTTCTGTGAGCATTTTGTGAACAAACAGAGCATCACCA 1458
 Qy 280 LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCCTTCTATTAATAAGCGGCACTGATTCGCGGTGCGAGCTGACATCGGCTTGGCTAC 1518
 Qy 300 ProSerGlyAspGlnGlyTyrGlyValArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1519 CCGACGCTAACCAAGATGGGACGAGTGACATGATGATTAATCCCTGAACCTGGCTAT 1578
 Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGAGTCCAGTCTCTATCCACCACGCCAAAGCGAGCTACTCGTTACTGCTACT 1638
 Qy 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
 Db 1639 GCGGCAAGCTTTGAAATATCTCCGTATGGTCTGATGCCCTCGACACCAATGCT 1698
 Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1699 TCCGTAACTGTCTCAATGATCTGACCTTCTCATACCGCTCCCAATGGCACACAGTAT 1758
 Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATTAATGAGTGGGTGGCCCAATACGTAGAA 1818
 Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGTATTATTATATGACCAACGAGCGGACGATATCAATGAGTACAGGCTTATAAC 1878
 Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGTGTGGACCAAGAACTTCTCGTTGCAATTTGGAAT 1920
 RESULT 12
 AX839476
 LOCUS
 DEFINITION
 Sequence 2 from Patent EP1347044.
 ACCESSION
 AX839476
 VERSION
 AX839476.1 GI:39922766
 KEYWORDS
 Bacillus sp. KSM-KP43
 SOURCE
 Bacillus sp. KSM-KP43
 ORGANISM
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE
 1

AUTHORS
 TITLE
 JOURNAL
 PATENT: EP 1347044-A 2 24-SEP-2003;
 Kao Corporation (JP)
 FEATURES
 source

1. .1305
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="unassigned DNA"
 /db_xref="taxon:109322"
 <1..1305
 /notes="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CA855146.1"
 /db_xref="GI:39922767"
 /translations="NDVARGIVKADVAOSSYGLYGQGIIVAVATGLDTRNDSSMHE
 AFGRKITALVAGRTNNDVNGHGVTHVAGSLVNGSTNGKMAPANLVFOSIMDSGG
 GLGLSPNLQTLFQSAVSAGARLHFNHSCAAVNGAYTTDSRVDDYVKNMTILPAA
 GNGPGVGTISAPGTAKNAITVGAATENLPPSGSVADNINHVAFSSRGPTKGRKIP
 DVNAPGTTILSARSSLPASSFWANHDSKIAMGTSMATPIVAGNVAQLREHFVNR
 GITPKSLKAALISAGADIGLIGYPNGNGWRVTLDKSLNVAYNNESSLSOKAT
 YSTATAGKPLKISLVSDAPASTTASVTLVNDLVLIVAPNGTOYVGNDETSPYNDN
 WDRNNVENVFINAPQSTYTTIEVQAYNVFVPGQTFSLAVN"

CDS

Alignment Scores:
 Pred. No.:
 Score:
 Percent Similarity:
 Best Local Similarity:
 Query Match:
 DB:

2e-117 Length: 1305
 1998.50 Matches: 381
 94.24% Conservative: 28
 87.79% Mismatches: 24
 88.39% Indels: 1
 6 Gaps: 1

ORIGIN

US-09-985-689A-5 (1-433) x AX839476 (1-1305)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 1 AATGATGTTCGGGTGGAATGTCAAAGCGGATGTGGCTCAGACGACGTACGGGTGTAT 60
 Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAAGACAGATCGTAGCGGTTCGATACAGGCTTGATACAGGTCCCAATCAGCT 120
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCGATGATGAAGCCTTCGCGGGAAATTAATCTGCAATATATGCAATGGGACGACGAAT 180
 Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 181 AATGCCAATGATACGAATGCTCATGTGTACCATGTGGCTCGCTATTAGGAAACGGC 240
 Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 241 TCCACTAATAAGAAATGGCGCCTCAGGCGAATCTAGTCTTCAATCTATCATGATAGC 300
 Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
 Db 301 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAACCTTATTCAGCCAACATACACT 360
 Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValaTyrThrAla 139
 Db 361 GCTGTCGCAATTCATACAAACTCTGGGAGCAGCAGTAGTAATGGGCTTACACAACA 420
 Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 421 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATCAGCATCTTTCGCTGCC 480
 Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 481 GGGAAATGAGACCGAAGCGGGAACCATCATGTCGACAGGACAGCAGCTAAAATGCCAATA 540
 Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199

```
Db 541 ACAGTCGAGCTACGGAAACCTCGCCCAACCTTTGGTCTTATGCGGCAATATCAAC 600
Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTGGCACAGCTTCTTTCAGTGGACCGCAAAAGGATGGAGGATCAAAACCGGATGTC 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCACCGGAACTTCTATCTATCAGCAAGATCTTCTCTGACCGGATCTCTCTTC 720
Qy 240 TrpAlaAsnTyrAsnSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db 721 TGGCGCAACCATGACAGTAATATGATACATAGGTTGGAACGTCCTCGGTACACCGATC 780
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTTGTGGAAACGTGGACAGCTTCTGTAGCATTTTGTGAGAAAACAGAGGATCACACA 840
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValcIleuGlyTyr 299
Db 841 AAGCCTTCTCTATTAAGACGCGCACTGATTGCGGTGCGAGTGCATCGGCCCTTGCTTAC 900
Qy 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerIleuAsnValAlaTyr 319
Db 901 CCGAACGGTAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGCCCTAT 960
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 961 GTGAACGAGTCCAGTTCTCTATCCACCGCAAAAGCGACGTACTCGTTTACTGCTACT 1020
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1021 GCGCGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCTGCGAGCACAACTGCT 1080
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCGGTAAAGCTTGTCAATGATCGGACCTTGTCATTCGCTCCAAATGCGCACACAGTAT 1140
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTAGGAAATGACTTTACTTCGCCATACATGATAACTGGGATGGCGCGCAATACGTAGAA 1200
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1201 AATGTATTTATTATGACCAACAAAGCGGACGTATACAAATTGAGGTACAGGCTTATAAC 1260
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTACCGGTTGGACACACAGACCTTCTCGTTGGCAATTTGTAAT 1302

RESULT 13
AB051423
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
FEATURES
source
gene
CDS
Location/Qualifiers
1..1923
/organism="Bacillus sp. KSM-KP43"
/mol_type="genomic DNA"
/strain="KP43"
/db_xref="taxon:109322"
1..1923
/genes="PROF"
1..1923
/genes="PROF"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB55674.2"
/db_xref="GI:20521155"
/translacion="MRKKKVFSLVLSAAAILSTVALSNPSSAGAAFNFLDPKGIQTT
TDKSGKQGTGAAAPLVESENKLPKGLQKLETPANNKLHIIQFNGLPILEBTKQ
OLEKTKAKILDYIPDAVYIYVEGDKVKSATISIEHVESVEPLPIRIDPQLFTKGAS
ELYKVALDTOKKKEVQLEQIAQFALSNDVLYITAKPEVKMVDVARGIVKADV
AQSYYGLGGQIVAVADTGLDTRNDSSHEAPRGKITALVALGFTNNANDTNGHT
HVASGVILGNGSTNKGAPQANLVFQSIMSGGGLGGLPSNLQTLFSQAYSAGARHIN
SWGAAYNGAYTDSRVNDYVRKNDMTILFAAGNPGNGGTISAPGTAKNADTSVGHATE
NLSPSGSVADNINHVAFSSRGFTKDRIKPDVWAPGTFILSARSLAPADIGLGPEN
DSKRVYMGGTSMATPIVAGNVAOLREHFVNKRGITPKSLKALILAGADIGLGPEN
GNGGWRVTLDKSLNWAYVNSSSLTSOKATYSTFATAGKPLKISLVMSDAPASITA
SVTLVNDLILVITAPNGTQYVGNDFSPYNDNWDGRRNVWFVFINAPQSGTYTIEVQA
INVPVGPQTFSLAIVN"
Alignment Scores:
Pred. No.: 3..17e-117 Length: 1923
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.39% Indels: 1
DB: 1 Gaps: 1
US-09-985-689A-5 (1-433) x AB051423 (1-1923)
Qy 1 AsnAspValAlaIleArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATGTGGCGGTGGAAATGTCAAAGCGGATGTGGCTCAGACGAGTACGGGTGTAT 678
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAGGACAGATCGTAGCGTTGCCGATACAGGGCTTGATACAGGTGCGCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyValThrAsn 60
Db 739 TCGATGTCATGAAGCCTTCCCGCGGAAATTAATCTCATTTATATGCTTGGGACGACCAAT 798
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCATGATACGAATGGTTCATGTCAGCATGTGGCTGGCTCCGTATTAGGAACCGC 858
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTATAAAGGAATGCGCCTCAGCGCAATCTAGTCTTCCAATCTATCATGATAGC 918
Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 919 GGTGGGACCTTGGAGGACTACCTTCGATCTGCAACCTTATTTCAGCAAGCATACAGT 978
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValThrAla 139
Db 979 GCTGGTGCAGAAATTCATACAAACTCCTGGGAGCAGCAGTGAATGGGGCTTCACACA 1038
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 1039 GATTCCAGAAATGTCGATGATGTCGCGCAAAATGATATGACGATCTCTTTTCGCTGC 1098
Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
```


Db 1099 GGGAATGAAGGACCGAAGCGGGAACCATCATGTCACCGACAGCTAATAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199

Db 1159 ACAGTCGGAGCTACCGAAGAACTCCGCCAAGCTTTGGGTCTTATGGGCAATATCAAC 1218

Qy 200 HisLeuAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219

Db 1219 CATGTGGCACAGTCTCTCTTCACGTGGACCGAAGCATGACCGATCAACCGGATGC 1278

Qy 220 ThrAlaProGlyThrPheIleuSerAlaArgSerSerLeuAlaProAspSerPhe 239

Db 1279 ATGGCACCGGAGCTTCATCTATCACAAGATCTCTCTCTGACCGGATCTCTCTCTC 1338

Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259

Db 1339 TGGGCGAAGCATGACAGTAAATATGCATACATGGGTGGAACGCTCCATGGCTACACCGATC 1398

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279

Db 1399 GTTGTGGAAAGCTGGACAGCTTCGTGAGCATTTTGTGAAAACAGAGGATCACACCA 1458

Qy 280 LysProSerLeuIleLysAlaLeuAlaLeuAlaGlyAlaThrAspValGlyLeuGlyTyr 299

Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCAGCTGACATCGCGCTTGGCTAC 1518

Qy 300 ProSerGlyAspGlnGlyTyrGlyValThrLeuAspLysSerLeuAsnValAlaTyr 319

Db 1519 CCGAAGCTTAACCAAGATGGGACGAGTGACATGTTGTAATATCCCTGAACGTTGCTAT 1578

Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339

Db 1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAAAGCGAGTACTCGTTTACTGCTACT 1638

Qy 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrAla 359

Db 1639 GCGGCAAGCTTTGAAATCTCCGTGTATGGTCTGATGCCCTGCGACCAACTGCT 1698

Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyLysTyr 379

Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTCTATACCGCTCCAAATGGCACACAGTAT 1758

Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyTyrAsnValGlu 399

Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATGATGATGATGATGATGATGATGATGAT 1818

Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419

Db 1819 AATGTATTATTAAATGCAACCAAGCGGACGATATACAAATGAGGTACAGGCTTATAAC 1878

Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433

Db 1879 GTACCGGTGGACACAGACCTTCTCTGTGGCAATTGTGAAT 1920

RESULT 14

AR368117
LOCUS 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 5 from patent US 6376227.
ACCESSION AR368117
VERSION AR368117.1 GI:34601778

KEYWORDS

Unknown.

SOURCE

Organism="unknown"

/mol_type="genomic DNA"

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 1923)

Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,

Shikata, S. and Nomura, M.

Alkaline protease

Patent: US 6376227-A 5 23-APR-2002;

Location/Qualifiers

1...1923

Source

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 3,17e-117 Length: 1923
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.39% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-5 (1-433) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20

Db 619 AATGATGTTGGGTGGGAATTTGCAAGCGGATGTGGCTCAGAGCAGCTACCGGTGTAT 678

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GGACAGGACAGATCGTAGCGGTGGCGATACAGGCTTGATACAGTGCATCAGCT 738

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 739 TCGATGATGAGCTTCCCGGGAAATTTACTGCATTATATGCATTTGGACGCGAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79

Db 799 AATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 858

Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99

Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAACTCATATGATGATGATGAT 918

Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119

Db 919 GGTGGGAGCTTGGAGGACTACCTTCGAATCTCCAAACCTTATTCAGCAAGCATACAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139

Db 979 GCTGGTGCAGAAATTCATACAACTCTCGGGGAGCAGCTGAATGGGCTTACACCA 1038

Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159

Db 1039 GATTCAGAAATGTGGATGACTATGTGGCGAAAAATGATATGATGATGATGATGATGATGAT 1098

Qy 160 GlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179

Db 1099 GGGAAATGAAGACCGAAGCGGACCATCAGTGCACCGACAGGACAGCTAAAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199

Db 1159 ACAGTGGAGCTACGGAAACCTCCGCGCAAGCTTTGGGTCTTATGCGGACAAATATCAAC 1218

Qy 200 HisLeuAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219

Db 1219 CATGTGGCACAGTCTCTCTTCAGTGGACCGAAGATGAGCGATCAACCGGATGTC 1278

Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239

Db 1279 ATGGCACCGGAGCTTCATCTATCACAAGATCTCTCTCTGACCGGATCTCTCTCTC 1338

Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259

Db 1339 TGGGCGAAGCATGACAGTAAATATGCATACATGGGTGGAACGTTCCATGCTACCGGATC 1398

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279

Db 1399 GTTGTGGAAAGCTGGACAGCTTCGTGAGCATTTTGTGAAAACAGAGGATCACACCA 1458

Qy 280 LysProSerLeuIleLysAlaLeuAlaLeuAlaGlyAlaThrAspValGlyLeuGlyTyr 299

Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCAGCTGACATCGCGCTTGGCTAC 1518

Qy 300 ProSerGlyAspGlnGlyTyrGlyValThrLeuAspLysSerLeuAsnValAlaTyr 319


```
/db_xref="GI:9664581"
/translation="MCGNIFLFSKRHPAQLIALISLLFVGTQMFIRSHAAGGGDIDL
TVTRQFENQAVYGLGILLFVYALITEVVRHTLAATLGLGLAVIALNHSYSEAL
SLKAVTMDNWTMLLTPTVKIAQVLDKPIPLISEVLPFSNIGGAATMIGDPPNI
TAVLSAFLDNVTMLLTPTVKIAQVLDKPIPLISEVLPFSNIGGAATMIGDPPNI
MIGASNPDAISGNKDPNIALASDGVTFNDRTIEMAPGILMTIYSPFMLIRWLYKDS
FGSRDRDTELRQVAIKDKIKLGVSGAILILVILGFLPHVILHAYSWALSALVY
MLLATHREHEEPELVEWETILLFPAFLFVLVHSLQMGVVIDYIEYVEKAIKFPDD
YRLAAALILWYSAIASAFIDINIPYATMIPVLSLAFELNPLNLIWALAPGCL
GGNGLTIGASANTVAGMBEAGYP:SFNEFFRAGFPVLLSTFVYFMYLIVYVGG
EGDALTWKIVLVIGISLIGITIAQYSGRAKSPAEALVDDDFDEIVSTIKSVYIKTGS
KGESESE"
complement(12019..12282)
/note="37Fl1#149; contains KH domain"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAF97186.1"
/db_xref="GI:9664583"
/translation="MSVPNHI IQAKDTSLPITMRIGKSLDTSVVVELEGOLSSRL
VKYKNGRLYPRELKTVNIMAEKTNSSVVFARGNVAFWRN"
complement(12279..12614)
/note="37Fl1#10; contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97185.1"
/db_xref="GI:9664582"
/translation="MGRSRLNRDKRKHAINWQTQLLSLLKNPSQFTPEQRSPFALH
LVKYSRRHRLPSEVREIMCRNTLLRYGENATIRFRNGHKIQTCHSCTSRRIPY
RQHSQKVK"
complement(12624..13274)
/note="37Fl1#11"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAF97187.1"
/db_xref="GI:9664584"
/translation="MDEGRQGLNIVLFPPGAGKGTQQRMMDATGLPQVSTGMDML
RAAKAGTSVLEAKVMDKALVDSVIDLRDLRDDAKNGVDFGPRTPVQA
BALSEIVEVSVAIDVDERIVERIGRSGCTGVVHTFNPIONGICGCGSTE
KRRADDEETVLAEAYHTQTSPLADFYRANGIFHHVDGDRDIDVITDLLAVLG"
complement(13284..16691)
/note="37Fl1#12"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97188.1"
/db_xref="GI:9664585"
/translation="MYMOLSGTSQATAVAGGSASLAREYLREVAGINKPSASLIKATL
INGAEDLTPIPNANEGWGOIDLESLNPSGSLVDFQDDERELOAGFSLIYSPD
LDGSGKIDITLWSDASANAQSESLNLLDLILAPDGSYLGNDFSSGISTTG
GSANLNMIERIRPAGATTONGDMVTVLHRGGSSORYSIVIAADATLPKADLTTF
GNSLPSSTELIGLVSISLAWNQGTLAGCPYRIQFEDITSGILYDGNRSLEGG
LDSVSFTQSTTGLRSLDNTNQNVPENLAINGINNIDIDLEIETIAQGLRVY
FQNGQSG:PTTSQDRDSAAVMDVRNETGLSLFPVIAHEGTGRPRVSLVSSVQSEP
PTYPILLSPEDSKSVNQTVFTISQGGENDTIYTLNLDTSASFDGATKRYAR
AGSFIVDYATYQNQPTVTSQRLYIEIGQVDSVDPVPSGTSVSXKPGERTGSIY
RNTGNSAQYTMSCSTSSOWOIMLGNSSSLEFEPLNLIQDLSMDVNVFPISTANGE
PLAGSDOVCVTSPTSLNRVTDVAVSELKAFPSDLYGPDGAVGPGALAPVY
ADTGLVFNHTIQNKGNVLPDPAVTLERGNPGWAELQFDEQISLSTVTLQPGOS
GVENRLLVPEPARNSNTYTLRVESFPQMTLNGSLVNGENLGNLNSVNGTLLA
APVNDFTFTEFVNSGNSDLLEWSTSLAPDNGSIGSNPPTSVPVLSQVQLAI
KAPNOTASGFTDNGMLFVNGTNRFTNAELLVIEVAATFAGISVDETDIAPLFI
PRGDSKQSIITFNGENIPLSGLTVEVDGEGNILDRTSSTISPSITDIAIGESIE
VVGKYSTDEDSIDGRMLVWLTITIDGVVIEFMADTSVSSQSGGIFGLIYLSYV
LVLALLGVLYGGRILKQSSKMDDDGTVELVADPAHTADHIGTAREQALDISHVNDI
ASGEVSQBIAAALAQSLDMPVPTKQAVPTGRPPSGLPSNGLPPVGLPPAGL
```

```
CDS
PPAKSVPLSVQVTAGPPLPPGLPDGWTMEQWNYHGEQYLQRMGLN"
complement(16790..18130)
/note="37Fl1#13"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97189.1"
/db_xref="GI:9664586"
Alignment Scores:
Pred. No.: 5,09e-28 Length: 60006
Score: 612.50 Matches: 166
Percent Similarity: 53.30% Conservative: 76
Best Local Similarity: 36.56% Mismatches: 157
Query Match: 27.09% Indels: 56
DB: 1 Gaps: 17
US-09-985-689A-5 (1-433) x AF268611 (1-60006)
QY 13 AlaGlnAsnAsnTyGlyLeuTyGlyGlnGlnValValAlaAlaAAsThrGly 32
Db 17393 GCTCCTCTCTTCTTGAATCTTGATGTTCTCGAGAAACGATTGCAGTGCATGCATGC 17324
QY 33 LeuAspThrGlyArgAsnAspSerSerMethHisGluAlaPheAArgGlyLeuThrAla 52
Db 17323 GTCGATATGGAT-----CATCTGACCTCATCGGACGGGTTCGACGA 17282
QY 53 LeuTyAla---LeuGlyArgThrAsnAlaAsnAspProAsn---GlyHisGlyThr 70
Db 17281 ATTATATCTCAGTTCGTTAGATAGTTCTCAATTGATAGCAACAGTGTGCATGGACG 17222
QY 71 HisValAlaGlySerValLeu-----GlyAsnAlaLeuAsnLysGlyMetAla 86
Db 17221 CACGTGGGTATGACTGTGCTTGGCGATGCTAGCGGTGATGCAGTAGCGACCGGTGTGCT 17162
QY 87 ProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyLeuGlyGlyLeu 106
Db 17161 CAGAAAGCAATCTTGTCTATGCTCTTGAGCAGCAGCCACCGGTGTCTTTGGAAGG 17102
QY 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAArgIleHisThr 126
Db 17101 CAAGGTTCAATCTATGATCTCTCTCTGATCGGAATTAACACAGCAAGTGTCCAGATC 17042
QY 127 AsnSerTrpGlyAlaProValAsn---GlyAlaTyThrAlaAsnSerArgGlnValAsp 145
Db 17041 AATGATGGGGTCTTAATGGAATTAACGAGCTACACTGCTGA-TCAAGATCGACAGAT 16983
QY 146 GluTyValArgAsnAsnAsp---MetThrValLeuPheAlaAlaGly---AsnGluGly 163
Db 16982 CAATATGCTCCACAAATAATATCTCTCTCTCTTTCGGTGTGGTATAATCGAGGC 16923
QY 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 16922 ACAGGTTCTTCAAGATTACTCTCGAGAACAGCAGCAAGATGTTCTTCCATTGCCTCA 16863
QY 184 ThrGluAsnTyArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
Db 16862 AGCGTTTAAT-----GTAGC-----GTACGCTCC 16839
QY 204 PheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
Db 16838 TATTTCATCGAAGGTCCAACTCTCGATGGAAGAAATAACCTGATTTAGTCCCTCCCGT 16779
QY 224 ThrPheIleLeuSerAlaArgSerLeuAlaProAsp----- 236
Db 16778 GACGGTATCTGTTCTCCAGCAGCAGAGGAAAGCAAGCTGTTGTTGTTTCACTCTCGGT 16719
QY 237 SerSerPheTrpAlaAsnTyAsnSerLysTyAlaTyMetClyGlyThrSerMetAla 256
Db 16718 ATTGGTACTCACTCAATGCTCGTTCAATGATATATGATATGCAACTCAGCGCAACTTCT 16659
```

```

QY 257 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGly 276
Db 16658 ACAGCAGTAGCAGGTTGTTTCTGCGACGCTGAATATCTGCGTGAAGTTGCTGGA 16599
QY 277 IleThr---ProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspVal 295
Db 16598 ATCAACAACACCATCTGCAAGCCTCATTAAGGCAACTCTCATCAACGGGGCAGAAAGATTTA 16539
QY 296 GlyLeu---GlyTyProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSer 314
Db 16538 GGCACACCGGACATTCCTAACCAACACGAGGATGGGGCAAAATTTGATCTCGAAAATAGC 16479
QY 315 LeuAsnValAla-----TyrValAsnGluAlaThrAlaLeu 326
Db 16478 CTGAATCTCTTCGCTGCGGCTTCTCTGGATGCTTTTCAGGATGATGAGCGTGAACCTT 16419
QY 327 AlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
Db 16418 CAGGCAGGATTCTCATTTGATTTATTCCTTTGACCTTGATGGATCAAGGGTATTGACATT 16359
QY 347 SerLeuValTyrThrAspAlaProGlySerThrThrAlaSerTyrThr-----Leu 363
Db 16358 ACCTTGGCATGGAGCGATCGAGAGCCAGTGCAGAAATGCTGCACAATCTGAATCAGTCTA 16299
QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAsp 383
Db 16298 TTGAACAATTTAGATTGATCTTAATTGCACCCGATGGGTCTTCATCTTGGTAATGAC 16239
QY 384 PheSerTyrProTyrAsp-----AsnAsnTyrAspGlyArgAsnAsnValGluAsn 400
Db 16238 TTCTCTCTGGGATTTCACAACAGCGGGTTCGCTGATAATCTGAACAATATCGAGCGC 16179
QY 401 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrIleGluValGlnAla 417
Db 16178 ATTGGATCCCTCGAGGACGACAGACAGACGCGGTGATGGATGGTTACAGTTGACAT 16119
QY 418 TyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 16118 CGAGGAGGAAGTTCT-----CAGGCTACAGTATCGTTATT 16083

```

Search completed: April 4, 2004, 08:08:33
Job time : 3595.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 335.908 Seconds
(without alignments)
5476.111 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNYGLY.....EVQAYNVSPQGFSLAIVH 433

Scoring table:
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q/cor2_1/USPTO spec1/US0985689/runat_31032004_161806_4145/app query.fasta_1.3498
-DB=N_Geneseq_22Jan04 -QFMT=fastap -SUFFIX=ring -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0985689 @CN 1 1 1238 @runat_31032004_161806_4145 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2245	99.3	1299	2	AAQ27516
2	2044	90.4	3003	2	AAV82382
3	2002.5	88.6	1920	2	AAAX37277
4	1999.5	88.4	1923	2	AAAX37279
5	1998.5	88.4	1923	2	AAAX37278
6	443	19.6	1977	2	AAAT85667
7	443	19.6	1977	2	AAAT85926
8	422.5	18.7	1236	2	AAAX05920

9	422.5	18.7	1566	2	AAAT85668
10	422.5	18.7	1962	2	AAAT85695
11	422.5	18.7	1962	2	AAAX05929
12	401.5	17.8	1977	2	AAAT85669
13	373.5	16.5	2121	4	ABU54900
14	345.5	15.3	2539	2	AAAT61454
15	345.5	15.3	2809	2	AAAT61455
16	341.5	15.1	135638	7	ABX34289
17	334	14.8	1329	6	ABK74643
18	332	13.8	2532	2	AAQ29134
19	307	13.6	2835	2	AAAT08141
20	307	13.6	4765	2	AAAT08132
21	307	13.6	4765	2	AAAT85670
22	307	13.6	4765	2	AAAX05921
23	306.5	13.6	1859	2	AAAT85677
24	298.5	13.2	3743	9	ADD24901
25	295	13.0	1560	7	ABZ37569
26	285	13.0	59816	7	ABZ37516
27	285	13.0	59816	7	ABZ37515
28	283	12.5	898	2	AAAT08131
29	282	12.5	564	2	AAAT08134
30	282	12.5	564	2	AAAT85676
31	281.5	12.5	546	4	ABL53453
32	276.5	12.2	2166	6	ABU40498
33	276.5	12.2	2166	6	ABU40498
34	273	12.1	3788	9	ADD24905
35	270	11.9	807	2	AAQ11835
36	269	11.9	810	7	ABZ21470
37	269	11.9	866	2	AAQ30104
38	269	11.9	866	2	AAQ30146
39	269	11.9	866	2	AAQ31697
40	269	11.9	866	2	AAQ26410
41	269	11.9	1140	2	AAAT97502
42	269	11.9	1140	2	AAAT93806
43	269	11.9	1140	2	AAV69598
44	269	11.9	1140	3	AAZ59988
45	269	11.9	1140	3	AAAS9767

ALIGNMENTS

RESULT 1

AAQ27516
ID AAQ27516 standard; DNA; 1299 BP.

AC AAQ27516;

DT 05-FEB-1993 (first entry)

DE Alkali-protease Ya enzyme gene.

KW Alkali resistance; surface active agent resistance; detergency improver;

SS.

OS Bacillus sp. Y.

Key Location/Qualifiers

FT CDS 1..1299

FT /*tag= a

PN JP04197182-A.

PD 16-JUL-1992.

PF 28-NOV-1990; 90JP-00327110.

PR 28-NOV-1990; 90JP-00327110.

PA (LLOY) LION CORP.

XX WPI; 1992-288440/35.

DR P-PSDB; AAR26274.

PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 resistance and improves detergency.

PS Claim 3; Page 2; 17pp; Japanese.

CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 in alkali resistance and surface active agent resistance and improves
 detergency

SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.4e-168 Length: 1299
 Score: 2245.00 Matches: 429
 Percent Similarity: 99.77% Conservative: 3
 Best Local Similarity: 99.08% Mismatches: 1
 Query Match: 99.29% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-5 (1-433) x AAQ27516 (1-1299)

Qy 1 AsnAspValAlaArgGlyIleValIleAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 1 AATGATGATGACAGAGGATAGTAAAGCTGATGTCACAAAACAATACGGATATAT 60
 Qy 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAAAGTCACTAGTTCAGTACGACACAGGCTTAGATACAGGTCTGACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGTCATGAAGCATTCGCGGGGAAATCACAGCTCTTTACGCGTTAGGAACAATAAT 180
 Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAla 80
 Db 181 AATGCGAGTATCCGATGGCATGGCATGGCATGGCATGGCATGGCATGGCATGGCATGG 240
 Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
 Db 241 TTAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAAATCTATTATGGATAGCAG 300
 Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
 Db 301 GGAGGATTAAGTGGCTTACCATCGAATTAATACGTTTATTTAGTCAAGCTTCGAATGCT 360
 Qy 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 Db 361 GGAGCAAGAATTCATCTACTTCTGGGAGCCCGCAATTAATGGAGCGTACACTGCTAAC 420
 Qy 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 421 TCGAGACAAAGTGGATGAGTATGTTCCGAATAATCATATGACGGTACTTTTGGAGCTGGT 480
 Qy 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 481 AATGAAGTCTCTAATTCAGGAACAATTAGTCTCCAGGTACAGCGAAAAATGCTATACG 540
 Qy 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 Db 541 GTCCGCGCAACGGAATACTATCGCCCAAGCTTCGGTTCGATAGCAGATACCCCAATCAT 600
 Qy 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 601 ATTGCACAAATTTTCATCAGAGGAGGACTACGAGGAGATGGACGAATTAAGCCCTGACGTAACA 660
 Qy 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
 Db 661 GCTCTCGGACATTTATTTATCAGCACAGCTTCTTCCTTAGCTCCAGACTCTTCGTTTGG 720
 Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 Db 721 GCGAATTAACAGATAACGCGTATATGGGCGGTACCTCCATGGCGCACACCTATTGTT 780

Qy 261 AlaGlyAsnValAlaGlnLeuArgGlyHisPheIleLysAsnArgGlyIleThrProLys 280
 Db 781 GCGGGGAATGTCGCGCAATTAGTGGACATTTTATAAAAAATAGAGTATTACTCTCTAG 840
 Qy 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
 Db 841 CCTCTTTTATAAAGCTGCTTATCGCTGGTGTCTACTGATGTTGGTTTAGGATATCCT 900
 Qy 301 SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
 Db 901 ASTGGTGACCAAGCTGGGGGCGTGTACTCTAGATNAATCGTTAATGTAGCGTATGTC 960
 Qy 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
 Db 961 AATGAAGCAACTGCTCATTTAGCCAGACAAAAGCAAGTATTCGTTCCAGACCAAGCG 1020
 Qy 341 GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer 360
 Db 1021 GGTAAACCTTTAAATCTCGTTAGTATGACAGATGCTCTCTGGAAGTACAACTGCATCT 1080
 Qy 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
 Db 1081 TATACACTAGTTAATGATTAGTATCTAGTATCTGCTCCGAATGGACAAATATATGTA 1140
 Qy 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGluAsn 400
 Db 1141 GGAATGATTTTATGTTATCTTATGATAATACTGGATGGTGGCAGCAAACTGTCAGAAC 1200
 Qy 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
 Db 1201 GTATTATAAACCTCCGCAATCTGGAAGTATATAATTGAGGTTCAGGCTCAAGCGTATATGTA 1260
 Qy 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1261 CCATCTGCCCCACAGCGTTTCTCCTAGCTATCGTACAT 1299

RESULT 2
 AAV82382
 ID AAV82382 standard; DNA; 3003 BP.
 XX
 AC AAV82382;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE Bacillus JP170 protease gene.
 XX
 KW Protease; detergent; surfactant; leather processing; debittering;
 KW flavour; ss.
 XX
 OS Bacillus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 846..2771
 FT /*tag= a
 FT sig_peptide 846..944
 FT /*tag= b
 FT mat_peptide 1470..2768
 FT /*tag= c
 XX
 PN WO9856927-A2.
 XX
 PD 17-DEC-1998.
 XX
 PF 09-JUN-1998; 98WO-US012005.
 XX
 PR 12-JUN-1997; 97US-00873479.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 PI Sloma A, Christianson L;
 XX
 WI WPI; 1999-080908/07.

DR P-PSDB; AAW89547.
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX
 XX Claim 11; Page 52-53; 77pp; English.
 XX
 CC This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JPI170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JPI170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50),
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW82410-11), reconstruction of the 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins
 XX
 SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,2e-152 Length: 3003
 Score: 2044.00 Matches: 386
 Percent Similarity: 94.69% Conservative: 24
 Best Local Similarity: 89.15% Mismatches: 23
 Query Match: 90.40% Indels: 0
 DB: Gaps: 0

US-09-985-689a-5 (1-433) x AAW82382 (1-3003)

QY	1	AsnAspValAlaArgGlyValLeuValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr	20
DB	1470	AATGACGTGGCCCGTGGCATTGTAAGCAGACGTCGCACAAATAACTTTGGCTTATAT	1529
QY	21	GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyValArgAsnAspSer	40
DB	1530	GGACAAAGGACAGATTGTAGCATTTGCTGATATCGGCTTGATACAGGAAGAATGACAGT	1589
QY	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
DB	1590	TCGATGCATGAGCATTCGCGGTAGATTACCGCACATATATGCACTGGGCAGAACGAAT	1649
QY	61	AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyValAsnAla	80
DB	1650	AACGCCAATGATCCAAATGGACATGGACCCATGTTGCTGGATCTGTGTAGGAATGCT	1709
QY	81	LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer	100
DB	1710	ACAAATAAAGGATGGCAGCGCAAGCAATCTAGCTTTTCAATCTATTATGATAGTGGT	1769
QY	101	GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAla	120
DB	1770	GGAGGGCTGGGAGGACTACTGCTAATCTACAAACATTTATTCAGTCAAGCATATAGTGT	1829
QY	121	GlyValArgIleHisThrAsnSerTyrGlyValAlaProValAsnGlyValAlaTyrThrAlaAsn	140
DB	1830	GGAGCGAGAAATTCATACGAATTCATCGGGGGCTCCAGTAAACGGTCCCTATACGACAGAC	1889
QY	141	SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly	160
DB	1890	TCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATTTCTTTTCGGCGCGGA	1949
QY	161	AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr	180
DB	1950	AATGAGGGACCGGTACGGTACATCATGTCACAGGACAGACAAATAATGGATTACA	2009

QY	181	ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis	200
DB	2010	GTGGGGGACCGAAACCTAGCTCAAGCTTCGGATCTTATCGGATAAATTAACCAT	2069
QY	201	IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLeuProAspValThr	220
DB	2070	GTGTCTCAATCTCTTCACAGGTCCTACTAGATGGACGCTATTAAAGCGGACGTCATG	2129
QY	221	AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr	240
DB	2130	GCACGAGTACGATATATCTCTCTGCTAGATCATCATTTAGCTCCAGATCTCTCATCTGG	2189
QY	241	AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal	260
DB	2190	GCAACCACTAGATAGTAATAATATGCTACATGGGTGCTACTTCTATGGCTACTCAATTGTA	2249
QY	261	AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys	280
DB	2250	GCAGGTAATGTTGCCAATTAAGGGAGCATTTTGTGAAAAATAGAGGGGTACTCTTAAG	2309
QY	281	ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro	300
DB	2310	CCTTCCTTTTAAAGCTGCTTTAATTCAGGTCTGCGGATGTTGGACTTGGCTTTCCA	2369
QY	301	SerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal	320
DB	2370	AATGTTAACCAAGGATGGGGAAGATACGTTAGATAAATCCCTAATATGCGATTTTGTG	2429
QY	321	AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla	340
DB	2430	AATGAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCTGTTTACGGCTCAAGCT	2489
QY	341	GlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAlaSer	360
DB	2490	GGTAAACCCCTTAAAAATATCATCTGTTGCTCAGATGCCACAGGTAGCAGCAGGATCA	2549
QY	361	TyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal	380
DB	2550	CTAATCTTAGTGAATGATTTAGACTTAGTAATCACTGCACCAATGAATGAATACGTC	2609
QY	381	GlyAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsn	400
DB	2610	GGAAATGACTTTTACAGCACCGTATGATAACAATTTGGATGGCAGAAACACCTGGAAAT	2669
QY	401	ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal	420
DB	2670	GTGTATTATCAATGCTCTCAAGCGGAACGTATACAGTCAAGTGCAGGCTTACATGTA	2729
QY	421	ProSerGlyProGlnArgPheSerLeuAlaIleValHis	433
DB	2730	CCAGTAAGTCCGCAACCTTTTCTTTAGCGATTGTACAT	2788
RESULT 3			
ID	AA337277 standard; DNA; 1920 BP.		
XX	AA337277;		
XX	AC		
XX	XX		
DT	20-MAR-2003 (revised)		
DT	21-JUL-1999 (first entry)		
XX	XX		
DE	Bacillus alkaline protease encoding DNA.		
XX	XX		
KW	Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;		
KW	washing composition; oxidising agent; ss.		
XX	XX		
OS	Bacillus sp.		
XX	XX		
PN	WO9918218-A1.		
XX	XX		
PD	15-APR-1999.		
XX	XX		

PF 07-OCT-1998; 98WO-JP004528.
 XX 07-OCT-1997; 97JP-00274570.
 PR
 PA (KROS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI: 1999-287736/27.
 DR P-PSDB; AAY17087, AAY17089.
 XX Alkali protease from *Bacillus* used in washing powders.
 XX Disclosure; Page 53-58; 71pp; Japanese.
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-12; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,6e-149 Length: 1520
 Score: 2002.50 Matches: 383
 Percent Similarity: 94.24% Conservative: 26
 Best Local Similarity: 88.25% Mismatches: 24
 Query Match: 88.57% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-5 (1-433) x AAX37277 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 616 AATGATGTGCCAGAGGATTTGTCAAAGCGGATGTGGCACAGAGCAGCTACCGTTTGTAT 675
 QY 21 GlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 676 GGACAGGCCAGATGTGCGAGTGTGGGATGTGGATGTGGATGTGGATGTGGATGTGGAT 735
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 736 TCGATGATGAGGCTTCGCGGTAAATAAACAGCACTATATGCACTGGGTGGACCAAT 795
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 796 AATCGGAATGATACGAACGGTTCATGTPACCCATGTGGCAGGTTCGGTATTAGGAATGCG 855
 QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 856 GCAACGATTAAGAAATGGACCTCAAGCGAATCTGGTTTTTCAATCCATCATGTATAGC 915
 QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
 Db 916 AGTGGTGGCTTGGAGGCTTGCCTTCCAAATCTGCAACCTTATTCAGCCAAAGCATTCAGT 975
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaProValAsnGlyAlaTyrThrAla 139
 Db 976 GCAGTGGCAGAAATTCATCAAACTCTGGGGGCGACGGTGAATGGGGCTACACGACA 1035
 QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159

Db 1036 GATTCAGAAATGTGGATGACTATGTAAGGAAAAATGATATGATGACGATTTCTTTCGGCGCT 1095
 QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1096 GGGATGAAAGCCGAAACGGCGGTACCATCATGTCACCTGGTACGGCTAAAAACCCATA 1155
 QY 180 ThrValGlyValaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1156 ACAGTCGGCGCAACCGAAACCTGGGTCCAGCTTCGGTTCCTATGCAGATAATATTAAAC 1215
 QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1216 CACGTTGCACAGTTCTCTTCCGTCGCCGACAAAGATGGCGAATCAAGCTCATGTC 1275
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTCTTGCACCCGATTCCTCTTC 1335
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1336 TGGCGGAATCATGACAGCAAAATATGCTATATGGGTGGAAACGTCCATGTCACACCGATT 1395
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1396 GTTCCGGGGAATGTGCACAGCTCCCGTGGACATTTGTGAAATAATAGAGGATCATCTCT 1455
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1456 AAGCTTCCCTATTGAAGACAGCTTTGATTGCAGGTGCTGCTGATTTGGATTGGTTAT 1515
 QY 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1516 CCGAACGGAACCAAGGATGGGCGCGAGTACCCCTGGATAATTCGTTGAACGTGTGCTAT 1575
 QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1576 GTGACGAAATCCAGTGCCCTATCACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
 Db 1636 GCGGCAAGCCATTTGAAAATCTCCCTGTATGTGGATGCCCTGCCCTCAAGCATCTGCT 1695
 QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1696 TCTGTAAACCTGGTCAATGATTTGGATTGGTTCATTACAGCACCAACGAAACAGATAT 1755
 QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
 Db 1756 GTCGGAATCACTTCTCAGCACCATTTGACAATAACTGGGATGGCGCAATAACGTAGAA 1815
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1816 AATGTATTTTATTAATTCGCCCAAGTGGACATATATACCATTTGAGGTGCAAGCATATAAT 1875
 QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1876 GTGCGGTTGGACCAAAACTTCTCGTTGGCAATTTGTGAAC 1917

RESULT 4
 AAX37279
 ID AAX37279 standard; DNA; 1923 BP.
 XX
 AC AAX37279;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KM washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.

XX WO9918218-A1.
 XX PD 15-APR-1999.
 XX PF 07-OCT-1998; 98WO-JP004528.
 XX PR 07-OCT-1997; 97JP-00274570.
 XX PA (KAO) KAO CORP.
 XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 XX PI Shikata S, Nomura M;
 XX DR WPI; 1999-287736/27.
 XX PI P-PSDB; AAY17091.
 XX PI Alkali protease from *Bacillus* used in washing powders.
 XX PS Disclosure; Page 63-68; 71pp; Japanese.
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6,248-149 Length: 1923
 Score: 1999.50 Matches: 381
 Percent Similarity: 94.24% Conservative: 28
 Best Local Similarity: 87.79% Mismatches: 24
 Query Match: 88.43% Indels: 1
 Ds: 2 Gaps: 1

US-09-985-689A-5 (1-433) x AAX37279 (1-1923)

QY 1 AsnAspValAlaArgGlyVileValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 619 AATGATGTTGCGGTGGAAATTGTCAAAGCGGATGTGGCTCAGAGCAGCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGCAAGACAGATCGTACGGTTCGCCATACAGGGCTTGATACAGTTCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAGCCTTCGCGGGAAATTAATCTGCTATATATGATTGGGACGACGAT 798
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATGCCAATGATPACGAATGGTTCATGCTACGATGCGGTGGCTCCGTTATAGGAAACGGC 858
 QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTATTAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCATCTATCATGGATAGC 918
 QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCAAGCATACAGT 978
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139

Db 979 GCTGTGTCAGAAATTCATACAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 QY 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCTCTTTCTGCTGCC 1098
 QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGATGAAGACCGAACCGCGGAACCATCAGTCACCCAGGCACAGTAAAAAATGCAATA 1158
 QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACAGTCGGAGCTACGGAACCTCCGCCCAAGCTTTGGTCTTTATGCGGACAAATATCAAC 1218
 QY 200 HisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyValGlyIleLysProAspVal 219
 Db 1219 CATGTGGCACAGTTCTTTCAGTGGACCGACAAGGATGGACGGATCAAAACCGGATGTC 1278
 QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1279 ATGGCACCGGAAACGTTTATATCAGCAAGATCTTCTTTGCACCGGATTCCTCCTTC 1338
 QY 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGCGAACCATGACAGTAATATGATACATGCGGTGGACGCTCCATGGCTACCGGATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTGTCTGGAACCGTGGCACAGCTTCTGTAGCATTTTGTGAAAAACAGAGGCATCACACA 1458
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AGCCTTCTCTATTAAAGCGGACATGATGCGGTGCGAGCTGACATCGGCCCTTGGCTAC 1518
 QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
 Db 1519 CCGAACGGTAAACCAAGGATCGGACGAGTGACATGGATAAAATCCCTGAACGTTGCCTAT 1578
 QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGATGCCAGTTCTCTATCCACCGCAAAAGCGACGCTACTCTGTTTACTGCTACT 1638
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
 Db 1639 GCGCGCAAGCCTTGAAATCTCCCTGTGTATGCTGTGCTGCTGCGGACCAACTGCT 1698
 QY 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1699 TCGTTAACGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACAGTAT 1758
 QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATGATNACTGGGATGGCGCAATACCGTAGAA 1818
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGTATTTAATTAATGCACCAAGCGGACGATACAAATGGAAGTACAGGCTTATAAC 1878
 QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGGTGGACCAAGCAACTCTCTGTTGGCAATTGGAAT 1920

RESULT 5
 AAX37278
 ID AAX37278 standard; DNA; 1923 BP.
 XX AC AAX37278;
 XX DT 20-MAR-2003 (revised)
 XX DT 21-JUL-1999 (first entry)
 XX DE *Bacillus* alkaline protease encoding DNA.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX Bacillus sp.
 OS WO9918218-A1.
 PN 15-APR-1999.
 XX 07-OCT-1998; 98WO-JP004528.
 PF 07-OCT-1997; 97JP-00274570.
 PR (KAOS) KAO CORP.
 PA Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17088, AAY17090.
 XX Alkali protease from Bacillus used in washing powders.
 PT Disclosure; Page 58-63; 71pp; Japanese.
 PS The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 7 49e-149 Length: 1923
 Score: 1990.50 Matches: 381
 Percent Similarity: 94.24% Conservative: 28
 Best Local Similarity: 87.79% Mismatches: 24
 Query Match: 88.39% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-5 (1-433) x AAX37278 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValIleAlaAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATGTGCAAGCGGATGGCTCAGACAGCTACGGGTGTAT 678

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAAGACAGATCGTACCGGTTGCCGATACAGGGTTGNTACAGTCCGATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCCTCCCGGGGAAATTAAGTCAATATATGATGCAATGGGACGAGCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 799 AATGCCAATGATACGAATGGTCAATGGTACGATGGTGGCTGGCTCCGATTAGGAACGGC 858

Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAATCTATCATGGATAGC 918

Qy 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
 Db 919 GGTGGGGAGACTTGGAGGACTTACTTTCGAATCTGCAAACTTATTCAGCCAGCATACAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyValTyrThrAla 139
 Db 979 GCTGGTGCAGAAATTCATCAAACTCCTGGGAGACGAGTGAATGGGGCTTACACAA 1038

Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTGGCTGCG 1098

Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGAAATGAGGACCGAAGCGGAAACCATCAGTGCACCAAGGACACACTTAAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
 Db 1159 ACAGTGGAGTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGCGGACAATATCAAC 1218

Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
 Db 1219 CATGTGCACAGTTCTCTTCCAGTGGACCGACAAAGGATGGACGATCAAAACGGATGTC 1278

Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1279 ATGGACCGGGAAGCTTCATCTATCAGCAAGATCTTCTTGACCGGATTCCTCCTTC 1338

Qy 240 TrpAlaSerAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGCGGAACCATGACAGTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1398

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
 Db 1399 GTTCTGGAACGTTGGGACAGCTTCTGAGCATTTTGTGAAACACAGGACATCACACCA 1458

Qy 280 LysProSerLeuIleLysAlaAlaLeuAlaGlyAlaThrAspValGlyLeuGlyTyr 299
 Db 1459 AAGCTTCTCTATTAAGGCGGCACTGATTCGCGGTGACATGACATCGGCTTGGCTAC 1518

Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuValAlaTyr 319
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGATGATGATGATGATGATGATGATGATGATGAT 1578

Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
 Db 1579 GTGAACGAGTCCAGTCTCTATCCACAGCCCAAAAGCGAGCTACTCGTTTACTGCTACT 1638

Qy 340 AlaGlyLysProLeuLysIleSerLeuValThrThrAspAlaProGlySerThrAla 359
 Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGGACGACATGCT 1698

Qy 360 SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
 Db 1699 TCCGTAAACGTTGTCAATGATCTGACCTTGTCTTACCGCTCCAAATGGCACAGTAT 1758

Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnThrAspGlyArgAsnValGlu 399
 Db 1759 GTAGGAATGACTTTACTTCGCCATACATTAACCTGGATGGCGGCAATACGTAGAA 1818

Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
 Db 1819 AATGATTATTATTAAGTGCACCAACCAAGCGGACGTATACAAATGGAGTACAGCTATAAC 1878

Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGGTGGACCAACAGACCTTCTCGTTGGCAATTTGTAAT 1920

RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

XX AAX05926;
 AC
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE WO9856926 Seq ID 11.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX
 OS Thermococcus celer.
 XX
 PN WO9856926-A1.
 XX
 PD 17-DEC-1998.
 XX
 PF 04-JUN-1998; 98WO-JP002465.
 XX
 PR 10-JUN-1997; 97JP-00151969.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 PI WPI: 1999-080907/07.
 DR P-PSDB; AAW94840.
 XX
 XX Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Disclosure; Page 53-54; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,228-25 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.59% Indels: 68
 DB: 2 Gaps: 16
 US-09-985-689A-5 (1-433) x AAX05926 (1-1977)
 QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnValValAla 27
 DB 433 ATAGGGGCGGATACCGCTGTGGAACTCCCTCGGCTACGACGAAACGGTGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
 DB 493 ATCGTCGATACGGGTATAGACGGAAC-----CACCCGATCTCGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 DB 535 GGCAAGGTATAGGTGGTACGACCGCGCTCAACGCGAGGTGCGACCCCTACGTACGAC 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 DB 595 GGACACGGAAACCCACGTTGCGGGTATCGTTGCGGGAACCGGACGCGTTAACTCCCGATAC 654

QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 DB 655 ATAGCGTCCGCCCGCGGCGGAAGCTCGTCGGGTCAAGGTTCCTCGGTCCGACGCTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
 DB 715 GGAAGCTCTCCACCATCATCGGGGTGTGACTGGTGTGTCGAGAACAGACAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 DB 775 GGGATAAGGGTTCATCAACCTCTCCCTCGGTCTCTCCAGAGTCCGACGGAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValA-GAsnAsnAspMetThrValLeuPheAlaGly 160
 DB 835 CTCAGTCAGCCGCTCAACACGCTGGAGCGCGGTATAGTAGTCTCGGTCCGCCCGGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 895 AACAGCGGCGGACACCTACCGTGGCTCACCGCGCGCGCGAGCAAGTCTATAACC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 DB 955 GTCCGTGCA-----GTTGACAGCAACGACAAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 DB 982 ATCCCGAGCTTCTCCAGCGGGACCGACCGCGAGCGAAGCTCAAGCCGGAAGTCGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 DB 1042 GCGCCGCGGTGACATCATAGCCCGCGCGCCAGC-----GGAACAGCATGGGC 1092
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 DB 1093 ACCCGATAACAGCTACTACACCAAGCCCTCTGGAAACAGCATGGCCACCGGACGTT 1152
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
 DB 1153 TCGGGCGTGGCGCGCTATCTCCAG-----GCCACCGCGAGTGGACCCCGGAC 1203
 QY 280 -----LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
 DB 1204 AAGTGAAGACCGCCCTATCGAGACCGCGCATAGTCGCCCGCCAGGAGATGACGCGAC 1263
 QY 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
 DB 1264 ATCGCTAC-----GGTGGGTAGGTGAACGTCTACAAGGCC-----Lys 331
 QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
 DB 1303 ATCAAGTACGACGACTACGCGCAAGCTCACTTCCCGGTCTCGTCCGCCACAGGGAAGC 1362
 QY 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThr 351
 DB 1363 GCCACCCACACTTCGACGTCAGCGCGCCACCTTGTGACCGCCACCTCTACTGGAC 1422
 QY 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
 DB 1423 -----ACGGGTCTGAGCGACATCGACTCTTACCTTACCTC 1452
 QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 DB 1453 TACGACCCCAAC-----GGGAACGAGTTGACTTCTCTACACCGCTAC 1497
 QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 DB 1498 TAC-----GGCTTCGAGAGGTGGTCTACTACACCGCGCGGACCTGG 1545
 QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 DB 1546 ACGTCAAGTTCGTGACGTAC-----AAGGGCGCGCGCACTACCAAGTTCGACGTC 1596

QY 432 Val 432
 Db 1597 GTC 1599

RESULT 8
 AAX05920
 ID AAX05920 standard; DNA; 1236 BP.
 XX AC AAX05920;
 XX DT 06-MAY-1999 (first entry)
 XX DE Hyperthermostable protease fragment encoding DNA.
 XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX OS Pyrococcus furiosus.
 XX PN W09856926-A1.
 XX PD 17-DEC-1998.
 XX PF 04-JUN-1998; 98WO-JP002465.
 XX PR 10-JUN-1997; 97JP-00151969.
 XX PA (TAKI) TAKARA SHUZO CO LTD.
 XX PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 DR WPI; 1999-080907/07.
 DR P-PSDB; AAW94836.
 XX PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 XX PT gene encoding it, for large scale production of the protease for
 XX PT industrial use.
 XX PS Claim 6; Page 37-38; 82pp; Japanese.
 XX CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,89e-24 Length: 1236
 Score: 422.50 Matches: 138
 Percent Similarity: 43.96% Conservative: 55
 Best Local Similarity: 31.44% Mismatches: 167
 Query Match: 18.69% Indels: 79
 DB: 2 Gaps: 17

US-09-985-689a-5 (1-433) x AAX05920 (1-1236)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 58 AACTGGGATATGATGTTCTGGAATCAATAGGAATAATTCACACTGGAATTGAC--- 114
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyGlySylleThrAlaLeuTyrAla 55
 Db 115 -----GCTTCTCCAGATCTCCAGGAAGAAGTA----- 144

QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 145 ATTTGGTGGGTAGATTTTGTCAATGGTAGAGTTATCCATAGCATGACCATGACATGGA 204
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 205 ACTCATGTAGCTTCAATAGCAGCTGCTACTGGAGCAGCAAGTAATGGCAAGTACAAGGA 264
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 265 ATGGCTCCAGGAGCTAAGCTGGCGGAAATTAAGTTCTAGGTCCCGATGTTCTGGAAGC 324
 QY 103 LeuGlyGlyLeuProSerAsnLeuThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 325 ATATCTACTATATAATTAAGGGAGTTGAGTGGCGGTTGATACCAAGATAAGTACGGAAT 384
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 385 AAGTCATTAAATCTTCTCTGTTTCAAGCCAGAGCTCAGATGGTACTCAGCGCTCTAAGT 444
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 445 CAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGTTGCTGGCTGGAAACAGT 504
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 505 GGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTATTATCAGTTGGA 564
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
 Db 565 GCGTTGCAAGTAT-----GATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 592 AGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCCTGAGTTGTGTCTCA 651
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 652 GGAACCTGGATAATTGCTCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 702
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCTCAGTAGTGTGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 763 ATTCCAGCCCTCTTGTCTCCAA-----GCACCCCGAGCTGGAGCTCCAGACAAAGTA 813
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 814 AAAACAGCCCTCAGAAACTGCTGATATCGTAAAGCCAGATGAAATAGCCGATATAGCC 873
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 874 TAC-----GGTGCAGTAGGTTAATGCATACAGGCTATAAAC----- 912
 QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
 Db 913 TACGATAACTATGAAAGCTAGTGTCTCAGTGGATATCTGCCAACAAAGGCGACCAACT 972
 QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 973 CACCAGTTCGTTATTAGCGGAGCTTCGTTCTGTAAGTCTCCACATTATATATGGAATGCC 1032
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1033 AAT-----AGCAGCCTTGATCTTACCTCTACGAT 1062
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
 Db 1063 CCCAATGAAACCAGGTT-----GACTACTCTTAC-----ACC 1095
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413

Db 1096 GCTACTATGGATTGCAAGAGTTGGTTATTACACCCCACTGATGGACATGGCAATT 1155
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1156 AAGTTTGAAGTAC-----AGCGAAGTGCAAACTATCAAGTAGATGGTGA 1203

RESULT 9

ID AAT85668 standard; DNA; 1566 BP.
 AC AAT85668;
 XX
 DT 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX
 DE Pyrococcus furiosus protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
 OS Pyrococcus furiosus; DSM-3638.
 XX

XX Key Location/Qualifiers
 FH CDS 1..1566
 FT /tag= a
 FT /transl_except= (pos: 1282..1284, aa: Xaa)
 FT /note= "Xaa= Gly, Val"

XX WO9721823-A1.
 XX 19-JUN-1997.
 XX 07-NOV-1996; 96WO-JP003253.
 XX 12-DEC-1995; 95JP-00323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 XX Teunasaawa S, Kato I;
 XX WPI; 1997-332794/30.
 XX P-PSDB; AAW24122.
 XX
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 XX Claim 7; Page 90-91; 159pp; Japanese.
 XX
 XX This sequence represents the coding sequence for the protease from
 CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 XX Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 3.86e-24 Length: 1566
 Score: 422.50 Matches: 138
 Percent Similarity: 43.96% Conservative: 55
 Best Local Similarity: 31.44% Mismatches: 167
 Query Match: 18.69% Indels: 79
 DB: 2 Gaps: 17

US-09-985-689A-5 (1-433) x AAT85668 (1-1566)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 58 AACTTGGGATATGATGTTCTTGAATCACAATAGGAATAATTGACACTGGATTGAC--- 114

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 115 -----GCTTCTCATCCAGATCTCCAAGGAAAGTA----- 144
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 145 ATTGGGTGGGTAGATTTTGTCAATGTGTAGAGTATCCATACATGACCATGGACATGGA 204
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCGAAGTACAAGGA 264
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 265 ATGGCTCCAGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTCCGATGTTCTGGAAGC 324
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 325 ATATCTACTATTAATTAAGGAGTTGAGTGGCGCTTGATAACAAGATAAGTACGGAAT 384
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 385 AAGTCTAATTAATCTCTCTCTGTTCAAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT 444
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 445 CAGGCTGTTAATGACGCTGGGATGCTGGATTAGTTGTTGTTGCTGCCGTGGAACAGT 504
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 505 GGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGA 564
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
 Db 565 GCGTTGACCAAGTAT-----GATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 592 AGCTTCTCAAGCAGAGGCCAACTGCAGACGCGAGGCTTAAGCCTGAGTTGTTGCTCCA 651
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 652 GGAACCTGGATAATTGTCGACAGCAAGT-----GGAACTAGCATGGGTCAACCA 702
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGCAACTCTCTCAGTAGTGGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 763 ATTGCAGCCCTCTTGTCTCCAA-----GCACCCCGAGCTGGACTCCAGACAAAGTA 813
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATATGCGATATAGCC 873
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuValAla 318
 Db 874 TAC-----GGTGCAGTAGGGTTAATGCATACAAAGGTATATAAC----- 912
 QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
 Db 913 TACGATAACTATGCAAGCTAGTGTTCACITGGATATGTGCCAACAAAGGCGAGCAAACT 972
 QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 973 CACCAGTTCTGTTATTAGCGGAGCTCGTTCGTAATCCACATTATATCTGGGACAAATGCC 1032
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1033 AAT-----AGCGACCTTGATCTTTACCTCTCTACGAT 1062
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393

QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
 Db 1459 CCCAATGGAAACAGGTT-----GACTACTCTTAC-----ACC 1491
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
 Db 1492 GCCTACTATGATTCCAAAGAGTTGGTTATTACAAACCACTGATGGACATGACAAATT 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 Db 1552 AAGGTGTGAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGGTA 1599

RESULT 12

AAT85669

ID AAT85669 standard; DNA; 1977 BP.

XX

AC AAT85669;

DT 20-APR-1998 (first entry)

XX

DE Protease coding sequence.

XX

KW Protease; research reagent; thermal stability; ss.

XX

OS Synthetic.

XX

PN WO9721823-A1.

XX

PD 19-JUN-1997.

XX

PF 07-NOV-1996; 96WO-JP003253.

XX

PR 12-DEC-1995; 95JP-00323285.

XX

PA (TAKI) TAKARA SHUZO CO LTD.

XX

PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

PI Tsunawasa S, Kato I;

XX

DR WPI; 1997-332794/30.

XX

DR P-PSDB; AAW24123.

XX

PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT Industrially and as research reagents.

XX

PS Claim 11; Page 95-97; 159pp; Japanese.

XX

CC This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries

XX

SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2,368-22	Length:	1977
Score:	401.50	Matches:	130
Percent Similarity:	43.96%	Conservatives:	63
Best Local Similarity:	29.61%	Mismatches:	167
Query Match:	17.76%	Indels:	79
DB:	2	Gaps:	17

US-09-985-689A-5 (1-433) x AAT85669 (1-1977)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTTGGGATATGATGGTCTTGGAAATCAATAGGAATAATTGACACTGGAAATTGAC--- 510
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 540

QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGGTGGGTAGATTTCATGGTAGAGTTATCCATAGATACCATGACATCGAATGGA 600
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
 Db 601 ACTCATGTAGTTCATAGCAGCTGCTACTGGAGCAGCAAGTAATGGCAAGTACAAGGGA 660
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
 Db 661 ATGGCTCCAGAGCTAAGCTGGCGGAATTAAGTTCTAGGTGCCGATGGTCTCTGGAAGC 720
 QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
 Db 721 ATATCTACTATATTAAGGAGTTGAGTGGCGGTGATACAAAGATAAGTACGGAATT 780
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
 Db 781 AAGGTCAATTAATCTTCTCTGTTTCAAGCCAGAGCTCCGACGGAACCGACTCCCTCACT 840
 QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGCGCTCAACACCGCTGGAGCGGTATAGTAGTCTGCGTCGCCGCGGCAACAGC 900
 QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGGCCGACACCTACACCTGCGCTCACCCTCGCGCGAGCAAGGTCTATAACCGTGGT 960
 QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
 Db 961 GCA-----GTTGACAGCAACGACACATCGCC 987
 QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleGlyProAspValThrAlaPro 222
 Db 988 AGCTTCTCCAGAGGAGCGACCGCGGACGGAAGGTCTAAGCGGAAGTCTGTCGCGCCC 1047
 QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GCGGTGACATCATAGCCCGCGCGCGCAGC-----GGACACGAGTGGCCACCCCG 1098
 QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATAACGACTACTACACCAAGGCTCTGGAACCCAGCATGCCACCCGCGACGTTTCGGGC 1158
 QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
 Db 1159 GTTGGCGCGCTCATCTCCAG-----GCCACCCGAGTGGACCCCGGCAAGGTG 1209
 QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
 Db 1210 AAGACCCCTCATCGAGACCGCGACATAGTCCGCCCAAGGATAGCGGACATCGCC 1269
 QY 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
 Db 1270 TAC-----GGTGGGTAGGTGGAACGTCTACAAGGCC-----ATCAAG 1308
 QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----LysAlaThr 333
 Db 1309 TAGACGACTACGCCAAGCTCCTTCACCGGTCCGTCGCCACAGGAAGCGCCACC 1368
 QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
 Db 1369 CACACCTTCGAGTCAGCGCGCCACCTTCGTCACCGCCACCTCTACTGGGAC----- 1422
 QY 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 Db 1423 -----ACGGGTCGAGGACATCGACTCTACCTCTCTAGGAC 1458
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnTrpAsp 393
 Db 1459 CCCAAC-----GGGAACGAGTTGACTACTCTCTACACCGGCTACTAC--- 1500
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413

```
Db 1501 -----GGCTTCAGAGGTCGGCTACTACACCCGACCGCGGAACCTGGACGGTC 1551
Qy 414 GluValGlnAlaTyrAsnValPProSerGlyProGlnArgPheSerLeuAlaVal 432
Db 1552 AAGTCTGTCAGCTAC-----AAGGCGCGCGGAACCTACGAGTCGACGTCGTC 1599

RESULT 13
ID ABL54900 standard; DNA; 2121 BP.
AC ABL54900;
XX
XX 11-SEP-2003 (revised)
XX 31-MAY-2002 (first entry)
XX
DE T. yonsei subtilisin-like serine protease coding sequence.
XX
XX Subtilisin-like serine protease; ss.
XX
XX Thermococcus yonseiensis.
XX
XX Key Location/Qualifiers
XX CDS 142..1779
XX /tag= a
XX /product= "subtilisin-like serine protease"
XX
XX KR2000072141-A.
XX
XX 05-DEC-2000.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX (KIMY/) KIM Y S.
XX
XX Chang HJ, Kim DH, Byun YR, Kim YS;
XX
XX WPI; 2001-298092/31.
XX P-PSDB; ABB09483.
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermococcus yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;

Alignment Scores:
Pred. No.: 4.24e-20 Length: 2121
Score: 373.50 Matches: 132
Percent Similarity: 45.85% Conservative: 56
Best Local Similarity: 32.20% Mismatches: 139
Query Match: 16.52% Indels: 83
DB: 4 Gaps: 18

US-09-985-689A-5 (1-433) x ABL54900 (1-2121)
Qy 6 GlyValLeuValAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlnVal 25
Db 445 GGAATCACAAA-----GCAGGAGTGATTTGGAGTCACAGGAAAAATATACA 495
Qy 26 ValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
Db 496 ATAGCAATTATTGACACAGGTATAGACGGAATCAGCTTCACTCTCA----- 543
Qy 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyValArgThrAsnAsnAlaAsn----- 63
Db
```

```
Db 544 ---GGTGGAAAAATA-----ATAGGATGGAAGACACTTTATCAACACAAA 585
Qy 64 -----AspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db 586 ACTACACATACGACGACCAATGCCATGGCACTACAGTACAGTATGCTGCGAGGTACA 645
Qy 76 ValLeuGlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 95
Db 646 GGTGCTGGAACAGTCCTTTACAAAGGGGTTGCTCTGATGCTTTGTTGGTAGGAATAAAA 705
Qy 96 IleMetAspSerSerGly-----GlyLeuGlyGlyLeu 106
Db 706 GTTTTAGATCAATGAAGCGGACGACATGAGCAGCTGTAACTGACGAGAAATGACTGGGCT 765
Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAlaArgIleHisThr 126
Db 766 GTTCAAAATAAAGATGTATACGGGAATCAAGTTATAAATTAAGCTCGGCACCTTCTACA 825
Qy 127 AsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGlu 146
Db 826 AGTTCTGATGGA-----ACTGACTCTACCTCATTAGCAGTGAATAGAGCAGTAGAT--- 876
Qy 147 TyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSer 166
Db 877 -----ACGGTATTGTAGTAGTTGTAGCAGCAGAACTCTGCGCCCTGCAAAA 924
Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186
Db 925 TACACCATAGGTCCTCCCTGGTCTGCGGAAAAAGCCATAACAGTCGCGCAGCAATGGCAGAT 984
Qy 187 TyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 985 -----GTAGGTGAACCTGGCTTTAAC-----CTTGGACGCTTTCCAGC 1023
Qy 207 ArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGlyThrPheIle 226
Db 1024 CGCGTCTCTACTGCTGACGGAAGAATAAAACCTGACATCGCGGCCAGGATATAATA 1083
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnTyrAsnSerLys 246
Db 1084 ACTCGCGCAAG-----GCGAATCTGTAAATGGA 1113
Qy 247 TyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 1114 TATGTAACATACAGCGGTACAAGCATGCAACACCTTTTGTAGCAGCACTGTTGCTCTT 1173
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAla 286
Db 1174 ATGCTTAAC-----GCTAATCCAAATCTCACTCCAAATGATGCA-----AAAAAT 1218
Qy 287 AlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrp 306
Db 1219 ATAATAATGCTCTACTGCAAAAAGCTGGGGGCTCCAAAGCAAAACGTTGACTATGTGCA 1278
Qy 307 GlyArgValThrLeuAspLysSerLeuAsnValAla----- 318
Db 1279 GGAAGATTGGATGGCTATGAAGCTATAGGGGTAGCAGTAATTTAGAGGAAATAATATT 1338
Qy 319 -----TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThr 333
Db 1339 GATGTACCAATCATTTATTTTCAGGTATCTGCCGGTCTCTCGTACAGTACT 1398
Qy 334 TyrSerPheGlnAla---GlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAsp 352
Db 1399 TGGACTTTTACGCAACAAATAACAAGTTATCCAAATGCAATAACATAATTAATTCCTGAC 1458
Qy 353 AlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThr 372
Db 1459 -----TGGGCAAAATTACAAT-----CCCGATTTGATATATATCTCTAT 1497
Qy 373 AlaPro---AsnGlyGlnLysTyrValGly 381
Db 1498 GATCCGGGCGCATGGAGGGAAGACCCCTGGG 1527
```

1055	Db	-----CATCCGGACCTGAAGGGCCGGGTGACCCGGTCCAGAAGACTTCACC	1099
58	Qy	ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu	77
1100	Db	GC CGCGCCCGCGCCGCGCACAAAGGTGGGCGCACCGGACCCACACCTCGCTCGATCGGGCGC	1159
78	Qy	GlyAsnAlaLeuAen-----LysGlyMetAlaProGlnAlaAsnLeuVal	92
1160	Db	GGCAGCGGCGCCGAGTCCAAAGGCGCATGACAGGGCGTGCACCCGCGCGCGGATGCTCTC	1219
93	Qy	PheGlnSerIleMetAspSerSerGlyLeuGlyGlyLeuProSerAsnLeuAenThr	112
1220	Db	AACGGCAAGGTCTCTGCAGCACTCC-----GGTTTCGGCAGCACTCCGGCATC	1267
113	Qy	LeuPheSerClnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly	130
1268	Db	CTCGCGCGCATGAGTGGCGCGCGCGCGAGCGCGCGCTGCTGCTCAACATGACCTCGGC	1327
131	Qy	Ala-----ProValAsnGlyAlaIleThrAlaAsnSerArgGln	143
1328	Db	GGCATGGACACCGGAGACCGACCCGCTGGAGCGCGG	1366
144	Qy	ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly	163
1367	Db	GTCGCAAAAGTGTCCCGCGAAGAGGCGTCTCTTCGCCATCTCGCGCGCGGCACGAGGGC	1426
164	Qy	ProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIleThrValGlyAla	183
1427	Db	CCGGAGTCG-----ATCGGTTCGCGCGGACGCGGACCGCGCTCATCCGTGGCGCC	1480
184	Qy	ThrGluAenTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln	203
1481	Db	-----GTGCAACCAAGACAAAGCTTCGCGCAC	1507
204	Qy	PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro	222
1508	Db	TTCTCTCCACCGGCCCGCCCTCGGCGACGGCGCCATCAAGCGGACGTCAACCGTCCC	1567
223	Qy	GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn	242
1568	Db	GGCGTGACATCACGCGCGCTCGCGGAGGCAACGACATCGCCAGGAGGTGCGGTGAG	1627
243	Qy	TyrAsnSerIleYsThrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly	262
1628	Db	GGACCGCGCGGTACATGACCATCTCCGCGACGCTCATGCGGACCCCGCACGTCGCGGGC	1687
263	Qy	AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer	282
1688	Db	CGCGGGCCCTCTCGAAGCAGCAG-----CACCCCGCAC	1720
283	Qy	LeuIleLysAlaAlaIleu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr	299
1721	Db	TGGACCTCCCGCAACTGAAGGCGGCGCTCACCGGTCTCCACCAAGGCGCGC---AAGTAC	1777
300	Qy	ProSerGlyAspGlnGlyTyrGlyArgGlyValThrLeuAspLysSerLeuAsnValAlaTyr	319
1778	Db	ACCCGTTCCGCGAGGGTTCGGCGCGGATCCAGGCGCAGACGAGCGGTCCACGACGACG	1837
320	Qy	ValAsnGluAlaThrAlaLeuAlaThrGly-----	329
1838	Db	ATCGCGACCGCGTCTCGGTGAGTTCGCGGTCCAGCAGTGGCGGCACACCGACGACGAG	1897
330	Qy	-----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle	346
1898	Db	CCGGETCACCAAGCAGCTGACTTACCGCACTCGGCACCCAGGACGTCACTGAGAGCTG	1957
347	Qy	SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal	364
1958	Db	ACGTGACGCCACCGACCCCAAGGCAAGCGCGCCCGCGGGCTCTTTCACGCTGGGC	2017
365	Qy	AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe	384
2018	Db	-----GCCACCAACGGTACCGTCCCGCGCGCGGC-----	2047

Qy 385 SerTyrProTyrAspAsnAenTrpAspGlyArg-----AsnAsnValGlu 399
 Db 2048 AGCGCTCCGTCACATGACCGCGGACCGCGCTCGGGCGGACGGTGGACGCGCGTAC 2107
 Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
 Db 2108 TCGCGTACGTGTCTGCC---ACGGCGCGGCGACACGCTCCGACGCGCGCGGTG 2164
 Qy 414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
 Db 2165 CAGCGGAGTCTGAGTCTGACGACGTCGTCGCGC 2201

RESULT 15

AAAT61455
 ID AAAT61455 standard; DNA; 2809 BP.

AC AAAT61455;

DT 17-OCT-2003 (revised)

DT 06-OCT-1997 (first entry)

XX DhpA-mel chimeric gene.

XX asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.

XX Streptomyces viridosporus.

OS Streptomyces antibioticus.

OS Chimeric.

PH Key Location/Qualifiers

FT CDS 338..2809

FT /*tag= a

FT mat_protein 338..2539

FT /*tag= b

FT /product= "DhpA protein product"

FT /note= "from S. viridosporus"

FT mat_protein 2540..2809

FT /*tag= c

FT /product= "Melanin"

FT /note= "from S. antibioticus"

XX WO9705243-A1.

XX 13-FEB-1997.

XX 30-JUL-1996; 96WO-JP002147.

XX 31-JUL-1995; 95JP-00212975.

XX 29-FEB-1996; 96JP-00067478.

XX (SAOC) MERCIAN CORP.

XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;

XX Ishihiki K, Yoshiohka T;

XX WPI; 1997-145682/13.

XX P-PSDB; AAW13668.

XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts

PT on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral

PT derivatives useful for synthesis of cardiovascular drugs.

XX Claim 5; Page 37-43; 78pp; Japanese.

XX This sequence is a fusion gene encoding Streptomyces viridosporus dhpA

CC gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-

CC dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA

CC enzyme allows the efficient conversion of 4-substituted-1,4-

CC dihydropyridine esters to chiral partially hydrolysed derivatives, for

CC use in the synthesis of cardiovascular drugs suitable for the treatment

CC of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003
 CC to standardise OS field)
 XX Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 9,83e-18 Length: 2809
 Score: 345.50 Matches: 133
 Percent Similarity: 41.06% Conservative: 53
 Best Local Similarity: 29.36% Mismatches: 169
 Query Match: 15.28% Indels: 98
 DB: 2 Gaps: 19

US-09-985-689A-5 (1-433) x AAAT61455 (1-2809)

Qy 18 GlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 998 GGTACGACGCAAGGCGGTGAAGATCGCGTCTCTGGACACCGCGTGTGACACGAGC--- 1054
 Qy 38 AsnAspSerMetHisGluAlaPheArgGlyValIleThrAlaLeuTyrAlaLeuGly 57
 Db 1055 -----CATCGGACCTGAGGGCGGGGTGACCGGTCCAGAGACTTCACC 1099
 Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 1100 GCCCGCGCGCGCGCGGCGGACAGGTGGCGCACCGCACCCACCGTGGCTCGATCGCGCG 1159
 Qy 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
 Db 1160 GGCACGGCGCGCCAGTCCAGGGCAAGTACAAGGGCGTGCAGCCCGCGCGCGGATCTC 1219
 Qy 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuGlyLeuProSerAsnLeuAsnThr 112
 Db 1220 AACGCAAGGTCTCTCGACGACTCC-----GGTTTCGGCGACGACTCCGGCATC 1267
 Qy 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
 Db 1268 CTCGCGCGCATGTGAGTGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1327
 Qy 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
 Db 1328 GGCATGGACACACCGGAGACCGCGCTGGAGGCGGCG----- 1366
 Qy 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlnGly 163
 Db 1367 GTCGACAAGCTGTCCGCGGAGAGGGGTCTCTGTTCCTCCATCGCGCGCGGCGGCGGCGG 1426
 Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
 Db 1427 CCGGAGTCC-----ATCGGTTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1480
 Qy 184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
 Db 1481 -----GTCGACGACAAAGGCAAGCTCGCGCAC 1507
 Qy 204 PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
 Db 1508 TTCTCTCTCACCGGCCCGCGCTCGGCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1567
 Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 242
 Db 1568 GCGGTGGACATCACGCGCGCTCGCGGAGGGGCAACGACATCGCGCGGAGGCGGTGGTAG 1627
 Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1628 GGACCGCGCGGTACATGACCATCTCCGCGACGTGATGGGCGACCGCGGCGGCGGCGGCG 1687
 Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
 Db 1688 GCGCGCGCGCTCTGACGACGAGCAG-----CACCCCGCAC 1720
 Qy 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299

```
Db 1721 TGGACCTCGCGCAACTGAGGGCGCGCTCACC GGCTCCACCAAGGGCGGC---AAGTAC 1777
Qy 300 ProSerGlyAspGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGGTTCGAGCAGGAGTTCGGCGCGATCCAGCCGACCAAGCGCTCCAGCAGACCGTG 1837
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
Db 1838 ATCCCGACCCGGTCTCGGTGAGTTCGGGTCCAGCAGTGGCGGCACACCGACGAGGAG 1897
Qy 330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
Db 1898 CCGGTCAACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTACCGCTGAAGCTG 1957
Qy 347 SerLeuValTTPThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
Db 1958 ACGTCGACCGCACCGACCCCAAGGCAAGGCGGCCCGGGGGCTTTCACGCTGGGC 2017
Qy 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
Db 2018 -----GCCACCAACCGTGACCGTCCCGCGGGCGGC----- 2047
Qy 385 SerTyrProTyrAspAsnAsnTTPAspGlyArg-----AsnAsnValGlu 399
Db 2048 AGCGCTCCGTCGACATACCGCCGACACCCGGTTCGGCGGCACGGTGGACGGCGCGGTAC 2107
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
Db 2108 TCGGGTACTGTGTGCGC---ACGGCGCGCGGCGAGACGGTCCGCAAGCGCGCGCGCGGTG 2164
Qy 414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
Db 2165 CAGCGCAGGTGAGTCTGACGACGTACCGTCCGGC 2201
```

Search completed: April 4, 2004, 01:12:23
Job time : 362.908 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.4512 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNVGLY.....EVQAYNVPSQGRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09985689/runat 31032004 161807 4184/app query.fasta_1.3498
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN 1.1.142 @runat 31032004 161807 4184 -NCFU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DISPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2044	90.4	3003	2	US-08-873-479-41
2	2002.5	88.6	1920	4	US-09-509-814A-3
3	1989.5	88.4	1923	4	US-09-509-814A-5
4	1998.5	88.4	1923	4	US-09-509-814A-5
5	443	19.6	1977	3	US-08-894-818B-2
6	443	19.6	1977	4	US-09-445-472-11
7	422.5	18.7	1236	4	US-09-445-472-2
8	422.5	18.7	1566	3	US-08-894-818B-34
9	422.5	18.7	1962	3	US-08-894-818B-34
10	422.5	18.7	1962	4	US-09-445-472-15
11	401.5	17.8	1977	3	US-08-894-818B-6
12	350.5	17.5	2539	3	US-09-000-016-3

13	350.5	15.5	2539	4	US-09-514-340-3	Sequence 3, Appli
14	345.5	15.3	2809	3	US-09-000-016-1	Sequence 1, Appli
15	345.5	15.3	2809	4	US-09-514-340-1	Sequence 1, Appli
16	312	13.8	2632	1	US-07-671-376C-4	Sequence 2, Appli
17	307	13.6	2835	1	US-08-750-532-6	Sequence 6, Appli
18	307	13.6	4765	1	US-08-750-532-8	Sequence 7, Appli
19	307	13.6	4765	3	US-08-894-818B-7	Sequence 7, Appli
20	307	13.6	4765	4	US-09-445-472-5	Sequence 15, Appli
21	306.5	13.6	1859	3	US-08-894-818B-15	Sequence 7, Appli
22	283	12.5	898	1	US-08-750-532-7	Sequence 11, Appli
23	282	12.5	564	1	US-08-750-532-11	Sequence 11, Appli
24	282	12.5	564	3	US-08-894-818B-14	Sequence 14, Appli
25	269	11.9	1140	1	US-08-322-677A-11	Sequence 11, Appli
26	269	11.9	1140	3	US-08-322-677A-11	Sequence 11, Appli
27	269	11.9	1140	3	US-08-898-218-11	Sequence 11, Appli
28	269	11.9	1140	3	US-08-848-793-11	Sequence 11, Appli
29	269	11.9	1140	4	US-09-445-270-6	Sequence 6, Appli
30	269	11.9	1140	4	US-08-322-678-11	Sequence 11, Appli
31	269	11.9	1143	3	US-08-269-050-1	Sequence 1, Appli
32	269	11.9	1143	6	5336611-1	Patent No. 5336611
33	269	11.9	1320	6	5217878-1	Patent No. 5217878
34	269	11.9	1350	3	US-09-049-867-1	Sequence 1, Appli
35	269	11.9	2280	1	US-07-661-378A-1	Sequence 1, Appli
36	269	11.9	2280	1	US-07-918-318-1	Sequence 1, Appli
37	269	11.9	2280	2	US-08-413-724-1	Sequence 1, Appli
38	269	11.9	2280	3	US-08-853-494-1	Sequence 1, Appli
39	269	11.9	10216	2	US-08-975-154-1	Sequence 1, Appli
40	268	11.9	807	1	US-07-706-691G-72	Sequence 72, Appli
41	268	11.9	807	1	US-08-254-021-72	Sequence 72, Appli
42	268	11.9	807	2	US-08-618-446-72	Sequence 72, Appli
43	268	11.9	807	3	US-08-980-135-72	Sequence 72, Appli
44	268	11.9	807	4	US-09-585-798-72	Sequence 72, Appli
45	268	11.9	1140	1	US-08-322-677A-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

Sequence 41, Application US/08873479
Patent No. 5891701
GENERAL INFORMATION:
APPLICANT: Sloma, Alan
APPLICANT: Lynne, Christianson
TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
TITLE OF INVENTION: Having Protease Activity
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agria, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 3003 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,638-194 Length: 3003
Score: 2044.00 Matches: 386
Percent Similarity: 94.63% Conservative: 24
Best Local Similarity: 89.15% Mismatches: 23
Query Match: 90.40% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-5 (1-433) x US-08-873-479-41 (1-3003)

QY 1 AsnAspValAlaArgGlyLeuValAlaAlaGlnAsnTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGCGCATTTGTAAGCAGACGTCGCACAAATAAATCTTTGGCTTAT 1529
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGCAGAGGACAGATTCTGACAGTTGCTGATCTGCGCTTGATACAGGAAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATCCATGAACATTCGCGCGTAAAGATTACCGCACTATATGCACTGGCGCAAGCAAT 1649
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1650 AAGCCCAATGATCCAAATGGACATGGACCCATGTTGCTGGATCTGTGTAGGAATGCT 1709
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db 1710 ACAATAAAGGATGGCCAGCCCAAGCAATCTAGTCTTTCAATCTATATGATAGTGT 1769
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
Db 1770 GAGGGCTGGAGAGACTACCTGCTATCTACAAACATTTATCAGTCAAGCATATAGTGT 1829
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
Db 1830 GGAGCGAGAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTCCCTATACGACAGAC 1889
QY 141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
Db 1890 TCTCGAAATGTTGATGATATGTCAGAAAATGATATGACGATCTTTTGGCGCGCA 1949
QY 161 AsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1950 AATGAGGGACCCAGGTAGCGGTACAAATCAGTGCACAGCAAGCAAGCAAAATGCGATTACA 2009
QY 181 ValGlyAlaThrGluAsnTyrAcqProSerPheGlySerLeuAlaAspAsnProAsnHis 200
Db 2010 GTTGGGGCAACGAAACCTACGTCACGCTTCGAACTTATGCGGATATATATTAACCAT 2069
QY 201 IleAlaGlnPheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
Db 2070 GTTGTCTCAATCTCTTCCAGGAGTCTTACTAGAGATGGAGGTATTAAGCCGCGACGTCATG 2129
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 2130 GCACCAGGTAGCTATATCTCTGCTAGATCATCATTAGCTCCAGATTCTCATTTCTGG 2189
QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 2190 GCAACCATGATAGTAAATATGCCTACATGGTGGTACTTCTATGGCTACTCCAAATGTA 2249
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
Db 2250 GCAGGTAATGTTGCAATTAAGGAGCATTTTGTGAAAAATAGGGGGTAATCCTCAAG 2309

```

```

QY 281 ProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrPro 300
Db 2310 CCTTCCCTTTTAAAGCTGCTTTAATTCAGGTGCTGGAGTGTGGACTTGGCTTTCCA 2369
QY 301 SerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrVal 320
Db 2370 AATGTTAACCAAGGATGGGAGAGTAAGCTTAGATAAATCCCTAAATGTCGCAATTTGTG 2429
QY 321 AsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAla 340
Db 2430 AATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489
QY 341 GlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer 360
Db 2490 GGTAAACCCCTTAAATAATATCATCTGTTTGGTCAGATGCACAGGTAGCACCGCATCA 2549
QY 361 TyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrVal 380
Db 2550 CTAACCTTTAGTAATGATTTAGACTTTAGTAATCACTGCACCAATGGAACTAAATACGTC 2609
QY 381 GlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
Db 2610 GGAAATGACCTTTACGACCCGTATGATACCAATTTGGATGGCAGAAACCAACGTGGAAAT 2669
QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnVal 420
Db 2670 GTGTTTATCAATGCTCCTCAAGCGGAACGTATACAGTCGAGTGCAGGCTTACAATGTA 2729
QY 421 ProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 2730 CCAGTAAGTCCGCAACCTTTTCTTTAGCGATTGTACAT 2768

```

```

RESULT 2
US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSURISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 1,178-190 Length: 1920
Score: 2002.50 Matches: 383
Percent Similarity: 94.24% Conservative: 26
Best Local Similarity: 88.25% Mismatches: 24
Query Match: 88.57% Indels: 1
DB: 4 Gaps: 1

US-09-985-689A-5 (1-433) x US-09-509-814A-3 (1-1920)

```


QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 616 AATGATGGCGAGAGGATTTGTCAAAGCGGATGGGCAGACAGCATCGGTTGTAT 675
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCAAGCCAGATTGTCGAGTTCGCCGATCTCGGATTCGATACAGGAAGAAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCGATGAAGCCCTTCGCGGTAAAAATAACAGCACTATATGCACTGGGTCGACCGAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGGATGATACGAAACGGTCATGGTACCCATGTGGCAGGTTCGGTATTAGGAATGGC 855
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAAGGAATGCACCTCAAGCGAATCTGGTTTTTCAATCCATCATGATGATGC 915
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db 916 AGTGGTGGCCTTGGAGGCTTGCCTTCCAAATCTGCAAACTTATTCAGGCCAAGCATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 976 GCAGGTGCCAGAAATTCATAACAACCTCTGGGGGGCAGCGGTGAATGGGGCCTACACGACA 1035
QY 140 AsnSerArgGlnValAspGlnTyrValArgAsnAsnAspMetThrValIlePheAlaAla 159
Db 1036 GATTCAGAAATGTGGATGACTATGTAGGAAATGATATGACGATCTTTTCGGGCT 1095
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1096 GGGATGAAAGCCGACGCGTACCACTAGTGCACCTGGTACGGCTAAAAACGCCATA 1155
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 1156 ACAGTCGGCGCAACCGAAACCTGGCTCCAGCTTCGGTTCCTATGCAGATAATTAATTAAC 1215
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 1216 CAGGTGACAGTTCTCTCCGTGGCCGACAAAGATGGCGGATCAAGCCTGATGTC 1275
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1276 ATGGCCCGCAGGACATACATTTTATCAGCAAGATCTTCTTTCACCCCGATTCTCTCCTTC 1335
QY 240 TrpAlaAsnTyrAsnSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1336 TGGCGAATCATGACAGCAAAATATGCCTATATGGTGGNACGTCCTATGGCAACCCGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 1396 GTTGGGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAATAATAGAGGAATCACTCCT 1455
QY 280 LysProSerLeuIleLysAlaAlaIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1456 AGCCCTTCCCTATTTGAAAGCAGCTTTGATTCAGGTCGCTGATTTGGATTGGGTTAT 1515
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1516 CCGAACGGAACCAACGAGTGGGCGGAGTGACCTGGATAAATCGTTGTAACGCTGCCTAT 1575
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 1576 GTGAACGGAATCCAGTCCTCATCACTAGCCAAAGCCGACATATACCTTTTACTCAACG 1635
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrThrAla 359
Db 1636 GCGGGGCAAGCCATGAAATCTCCCTGGTATGGTCCGATGCCCTGCAAGCACTACTGCT 1695

QY 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1696 TCTGTAAACCTGGTCAATGATTGGATTGGTCATTACAGCCCAACGCAACAGATAT 1755
QY 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1756 GTCGGGAATGACTTCTCAGCACCATTTGCAATAACTGGGATGGCGCAATTAACGTAGAA 1815
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1816 AATGTAATTATTAATTCGCCCCCAAGTGGACATATACCATTCAGGTGCAAGCATATAT 1875
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1876 GTCCCGGTTGGACACAAAACCTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 3
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUO
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-7
Alignment Scores:
Pred. No.: 2,34e-190 Length: 1923
Score: 1999.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.43% Indels: 1
DB: 4 Gaps: 1
US-09-985-689A-5 (1-433) x US-09-509-814A-7 (1-1923)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 619 AATGATTTCCGCTGGAAATTCCTCAAGCGGATTCGGCTCAGACAGCTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAAGACAGATCGTAGCGGTTCGGATACAGGGCTTCATACAGCTCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGATGAAGCCCTTCGCGGGAATTAATCTGCTATATATTCGATTGGACCGACGAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 799 AATGCCAATGATACGAATGGTTCATGGTACGCATGTGGCTCCGCTCCGATTAGGAAACGGC 858

```

QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
DB 859 TCCACTAATAAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAATCTATCATGGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
DB 919 GGTGGGGGACTTGAGAGACTACCTTCGATCTGCAACCTTATTCAGCCAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpAla 139
DB 979 GCTGGTGCAGAAATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
QY 140 AsnSerArgGlnValAspGluValArgAsnAsnAspMetThrValLeuPheAlaAla 159
DB 1039 GATTCAGAAATGTGAGTACTATGTGGCAAAATGATATGACGATCTTTTCGCTGCC 1098
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
DB 1099 GCGAATGAAGGACCGAAGCGCGAACCATCAGTGCACAGGACAGCAGTAAATAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnTrpArgProSerPheGlySerLeuAlaAspAsnProAsn 199
DB 1159 ACAGTGGAGCTACGGAACCTCCGCGCCCAAGCTTTGGGTCTTATCGGACATATCAAC 1218
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
DB 1219 CATGTGCACAGTCTCTTCACTGGGACCGACAAAGGATGGAGCGATCAACCGGATGTC 1278
QY 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
DB 1279 ATGGCACCGGAGCTTCACTACTATCAGCAAGATCTTCTTGCACCGGATCTTCTCTTC 1338
QY 240 TrpAlaAsnTrpAsnSerLysTrpAlaTrpMetGlyGlyThrSerMetAlaThrProIle 259
DB 1339 TGGCGGACCATGACAGTAAATATGATCATATGGTGGAAAGCTCCATGGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
DB 1399 GTTGTGGAACCGTGGCACAGCTTCGTGAGCATTTTGTGAANAACAGAGGATCAACACCA 1458
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaThrAspValGlyLeuGlyTrp 299
DB 1459 AAGCTTCTCTATTAAAGCGGCACTGATTCGCGGTGACGTGACATCGGCTTGGCTAC 1518
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTrp 319
DB 1519 CCGAACCGTAACCAAGGATGGGACGAGTGACATTTGGATATAATCCCTGAAACGTGGCTAT 1578
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTrpSerPheGlnAlaGln 339
DB 1579 GTGAACGAGTCCAGTCTCTATCCACGCAAAAGCGAGCTACTCGTTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThrAla 359
DB 1639 GCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTGGGAGCACAACTGCT 1698
QY 360 SerTrpThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTrp 379
DB 1699 TCCGTAACGCTTGTCAATGATCTGACCTTGTCTATCCGCTCCAAATGGCACACAGTAT 1758
QY 380 ValGlyAsnAspPheSerTrpProTrpAspAsnTrpAspGlyArgAsnAsnValGlu 399
DB 1759 GTAGGAATGACTTTTACCTCGCCATACAAATGATAACTGGGATGGCCGCAATAAGCTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTrpThrIleGluValGlnAlaTrpAsn 419
DB 1819 AATGTATTATTAAATGCACCAAAAGCGGAGCGTATACAAATTAAGTATACAGCGCTATAAC 1878
QY 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
DB 1879 GTACCGGTGGACCAAGAACTTCTCGTTGGCAATTGTGAAT 1920

```

```

RESULT 4
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, AKAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PCT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 2,94e-190 Length: 1923
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.39% Indels: 1
DB: 4 Gaps: 1

US-09-985-689A-5 (1-433) x US-09-509-814A-5 (1-1923)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTrpGlyLeuTrp 20
DB 619 AATGATGTTGGCGTGAATTTGTCAGGCGGATGTGCTCAGAGCAGCTACGGGTGTGAT 678
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGACAAAGGACAGATCGTAGCGGTTCCCGATACAGGGCTTGATACAGGTCCGATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTrpAlaLeuGlyArgThrAsn 60
DB 739 TCGATGTCATGAAGCCTTCGCGGGGAAATTTACTGCAATTTATGCAATTCATCATGGAGT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
DB 799 AATGCCAATGATACCAATGGTTCATGTCAGCATGTGGCTGGCTCGCTCGTATTAGGAACGCG 858
QY 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
DB 859 TCCACTAATAAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAATCTATCATGGATAGC 918
QY 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
DB 919 GGTGGGGGACTTGAGAGACTACCTTCGATCTGCAACCTTATTCAGCCAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpAla 139
DB 979 GCTGGTGCAGAAATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
QY 140 AsnSerArgGlnValAspGluTrpValArgAsnAsnAspMetThrValLeuPheAlaAla 159
DB 1039 GATTCAGAAATGTGAGTACTATGTGGCAAAATGATATGACGATCTTTTCGCTGCC 1098

```

160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle	179
1099	GGGAATGAAGNACCCGAACGGCGGAACCATCACTGCACCAGGCACACAGCTAAAAAATGCAATA	1158
180	ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn	199
1159	ACAGTCGGAGCTACCGAAACCTCCGCCCAACGCTTGGGTCTTATGCGGACAAATATCAAC	1218
200	HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleIysProAspVal	219
1219	CATGTGGCACAGTTCTTTCATCGTGGACCGACAAAGGATGGACGGATCAACCGGATGTC	1278
220	ThrAlaProGlyThrPheIleIeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	239
1279	ATGGCACCGGGAACGTTCACTATCAGCAAGATCTTCTTGCACCGGATTCCTCCCTTC	1338
240	TrpAlaAsnTyrAsnSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	259
1339	TGGCGGAACCATGACAGTAANAATGCATACATGGGTGGAACGTCCATGGCTACACCGATC	1398
260	ValAlaGlyAsnValAlaGlnIeuArgGluHisPheIleIysAsnNrgIyleThrPro	279
1399	GTGTGTGGAAACGTGGCAGACTTCGTGAGCATTTTGTGAAAAACAGAGCGCATCACACCA	1458
280	LysProSerLeuIleIysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr	299
1459	AAGCCTTCTTATTAAGAAGCGCACTGATGCGGTGCAGCTGACATCGGCTTGCTACT	1518
300	ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspIysSerIeuAsnValAlaTyr	319
1519	CCGAACGGTAACCAAGGATGGGACGAGTGACATTCGATAAATCCCTGGAACGTTGCTAT	1578
320	ValAsnGluAlaThrAlaLeuAlaThrGlyGlnIysAlaThrTyrSerPheGlnAlaGln	339
1579	GTGAACAGTCCAGTCTCTATCCACCGCAAAAGACGCTACTCGTTACTGCTACT	1638
340	AlaGlyLysProLeuIysIleSerLeuValTrpThrAspAlaProGlySerThrAla	359
1639	GCCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCCCTCGACACCACTGCT	1698
360	SerTyrThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnIysTyr	379
1699	TCCGTAAACGCTTGTCATATGATCTGGACCTTGTCATACCGCTCCAAATGGCACACAGTAT	1758
380	ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu	399
1759	GTAGGAAATGACTTTACTTCGCCCATACAATGATAACTGGGATGCCCAATAACGTAGAA	1818
400	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	419
1819	AATGTATTTTAAATGACCACAAAGCGGACGTTATTAATTTAGGTACAGGCTTATAC	1878
420	ValProSerGlyProGlnArgPheSerLeuAlaIleValHis	433
1879	GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTGTGAAT	1920

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822

; PAGE# NO. 0201022
: GENERAL INFORMATION:

/ GENERAL INFORMATION:
 / APPLICANT: TAKAKURA, Hikaru
 / APPLICANT: MORISHITA, Mio
 / APPLICANT: YAMAMOTO, Katsuhiko
 / APPLICANT: MITTA, Masanori
 / APPLICANT: ASADA, Kiyoza
 / APPLICANT: TSUNASAWA, Susumu
 / APPLICANT: KATO, Ikunoshin
 / TITLE OF INVENTION: HYPERTHERMOSENSITIVE
 / NUMBER OF SEQUENCES: 42
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Broadway and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/894,818B
 FILING DATE: 20-MAY-1998
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP96/03253
 FILING DATE: 07-NOV-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 323285/1995
 FILING DATE: 12-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Browdy, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: TAKAKURA-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3328
 FORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1977 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 894-818B-2

Alignment Scores:

Alignment Scores:						
Pred. No.:	1.4e-34	Length:	1977			
Score:	443.00	Matches:	137			
Percent Similarity:	46.03%	Conservative:	66			
Best Local Similarity:	31.07%	Mismatches:	170			
Query Match:	19.59%	Indels:	68			
Gap:	3	Gaps:	16			

US-08-085-6892-5 (1-433) x US-08-894-818B-2 (1-1977)

Qy	8	VallyAlaAspValAlaGlnAsnAsnThrGlyGlnGlyValValAla	27
	:::		
	:::		
	:::		
Db	433	ATAGGGCGGATACCGTCTGGAACTCCCTCGGCTACGACGGAAGCGGTGTGTGGTTGCC	492
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethHisGluAlaPheArg	47
	:::		
	:::		
	:::		
Db	493	ATCGTCGATACCGGTATAGACGGAAAC-----CACCCCGATCTCGAAG	534
Qy	48	GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn	66
	:::		
	:::		
	:::		
Db	535	GGCAAGTCAATAGCTGGTACGACGCCGTCAACGGCAGGTCCGCCCTACGATGACCAAG	594
Qy	67	GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn	82
	:::		
	:::		
	:::		
Db	595	GGACACGGAAACCAAGTTCGGGGTATCGTTCGGGAAACCGGCAGCGTTAACTCCCAAGTAC	654
Qy	83	LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer	100
	:::		
	:::		
	:::		
Db	655	ATAGGCGTTCGCCCGCGCGGAAGCTCGTCGGCGTCAAGGTTCTCGGTGCCGACGGTTCCG	714
Qy	101	GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla	120
	:::		
	:::		
	:::		
Db	715	GGAAAGCGTCTCCACCATCATCGCGGGTGTTCAGTGGGTCTGCCGAACAAGGCAAGTAC	774
Qy	121	GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaIleThrAlaAsn	140
	:::		
	:::		
	:::		
Db	775	GGCATAGGCTCATCAACCTCTCCCTCGGCTCTCCCGAGAGCTCCGACGGAACCGACTCC	834

141 SerArgGlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGly 160
 835 CTCAGTCAGCCGCTCAACACCCCTGGAGCGCGGTATAGTAGTGGTGGTGGCGGC 894
 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAsnAlaIleThr 180
 895 AACAGCGGGCCGGAACACCTACCGCTCCGCTCACCCTGGCGCGCGGCAAGAGTGCATACC 954
 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 955 GTCGGTGCA-----GTTGACAGCAACGACAC 981
 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 982 ATCGCCAGCTTCTCCAGCAGGGAGCGACCGCGGACCGAAGGCTCAACCGGAGTCTGTC 1041
 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 1042 GCCCGCGGCTTGACATCATAGCCCGCGCGCAGC-----GGACAGCATGGGC 1092
 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 1093 ACCCGATAAAGACTACTACCAAGAGGCTCTGGAAACCCAGCATGGCCACCGGACGCTT 1152
 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
 1153 TCGGGGCTTGGCGGCTCATCTCCAG-----GCCACCCGAGTGGACCCCGGAC 1203
 280 -----LysProSerLeuIleLysAlaLeuAlaGlyAlaThrAspValGly--- 296
 1204 AAGGTGAAGACCCCTCATCGAGACCGCGCATAGTCGCCGCCCAAGGAGATAGCGGAC 1263
 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
 1264 ATCGCTAC-----GGTGGGTTAGGTGAACGCTCTACAGGCC----- 1302
 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
 1303 ATCAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1362
 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThr 351
 1363 GCCACCCACACCTTCGAGCTCAGCGCGGCGCCCTTCGTCACCGCCACCTCTACTGGGAC 1422
 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuValIle 371
 1423 -----ACGGGCTCGAGGACGACATCGACCTCTACTCTAC 1452
 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
 1453 TAGACCCCAAC-----GGGACGAGGTTGACTACTCTCTACACCGCCTAC 1497
 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
 1498 TAC-----GGCTTCGAGAGTGGCTACTACACCGGACCGCGGACGACCTGG 1545
 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 1546 ACGGTCAAGGTCTGTCAGCTAC-----AAGGGCGCGCGCAACTACGAGTCTGACGTC 1596
 432 Val 432
 1597 GTC 1599

RESULT 6
 US-09-445-472-11
 ; Sequence 11, Application US/09445472
 ; Patent No. 6358726
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko

APPLICANT: ASADA, Kiyozo
 APPLICANT: KATO, Kunoshin
 TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 FILE REFERENCE: TAKAKURA=6
 CURRENT APPLICATION NUMBER: US/09/445,472
 CURRENT FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: 151969/1997
 PRIOR FILING DATE: 1997-06-10
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patent in version 3.0
 SEQ ID NO 11
 LENGTH: 1977
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE: Information: Synthetic
 OTHER INFORMATION: US-09-445-472-11

Alignment Scores:
 Pred. No.: 1,4e-34 Length: 1977
 Score: 443.00 Matches: 137
 Percent Similarity: 46.03% Conservative: 66
 Best Local Similarity: 31.07% Mismatches: 170
 Query Match: 19.59% Indels: 68
 Gaps: 16

US-09-985-689A-5 (1-433) x US-09-445-472-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValValAla 27
 Db 433 ATAGGGCCCATACCGTCTCGAACTCCCTCGGTACACGGAAGCGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
 Db 493 ATCGTCGATACGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnProAsn 66
 Db 535 GCGAAGTCTACGCTGGTACGCGCGTCAACGCGAGGTCGACCCCTACGATGACCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
 Db 595 GGACACGGAACCCACGTTGGGTATCGTTCGCGGAACCGGACGCGTAACTCCAGTAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
 Db 655 ATAGCGCTCGCCCGCGCGAGCTCTCGCGGTCAAGGTCTCTCGGTGCGCGGTTCCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
 Db 715 GGAAGCTCTCCACCATCATCGCGGTGTTGACTGGGTCTGCCAGAACAGGACAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 Db 775 GGGATAAGGTCATCAACCTCTCCCTCGGTCTCTCCAGAGCTCCGACGGAACCGACTCC 834
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
 Db 835 CTCAGTCAGCGGTCACACCCCTGGACCGCGGTATAGTCTCGGTGCGCGCGGC 894
 QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 895 AACAGCGGGCGGAACACCTACCGTCCGCTCACCCTGGCGGAGCAAGGTCATACAC 954
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 Db 955 GTCGGTGCA-----GTTGACAGCAACGACAC 981
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 982 ATCGCCAGCTTCTCCAGCAGGGAGCGACCGCGGACCGAAGGCTCAACCGGAGTCTGTC 1041
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240

```
Db 1042 GCGCCGCGTGTGACATCATAGCCCGCGCCGAC-----GGAACGACGATGGC 1092
Qy 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCGATAACGACTACTACACAGCGCTCTGGACGACGATGGCCACCGCGCGTT 1152
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
Db 1153 TCGGCGGTGGCGGCTCATCTCCAG-----GCCACCGCGAGTGGACCCCGGAC 1203
Qy 280 -----LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--- 296
Db 1204 AAGTGAAGACCGCCCTCATGACGCCCGACATAGTCGCCCGCAAGGATAGCGGAC 1263
Qy 297 LeuGlyTyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
Db 1264 ATCGGCTAC-----GGTGGGCTAGGTGAACGTCTACAGGCC----- 1302
Qy 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
Db 1303 ATCAAGTACGACGACTAGCAAGCTCACCTTACCGGCTCCGTCGCCGCAAGGGAAGC 1362
Qy 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThr 351
Db 1363 GCCACCCACACCTTCGACGTCAGCGGCCACCTTCGTGACCGCCACCTCTACTGGGAC 1422
Qy 352 AspAlaProGlySerThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle 371
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTC 1452
Qy 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
Db 1453 TAGCACCCCAAC-----GGGAACGAGGTGTACTCTCTACACGCGCTAC 1497
Qy 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
Db 1498 TAC-----GGCTTCGAGAAGGTGCGCTACTACAACCCCGACCGCGGACCTGG 1545
Qy 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
Db 1546 ACGGTCAAGTGTCTAGCTAC-----AAGGGCGGCGGCAACTACCAGGTGCGACGTC 1596
Qy 432 Val 432
Db 1597 GTC 1599
```

RESULT 7

```
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2
```

Alignment Scores:

```
Pred. No.: 7,65e-33 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 4 Gaps: 17

US-09-985-689A-5 (1-433) x US-09-445-472-2 (1-1236)

Qy 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTGGGATATGATGTTCTCGAATCAATAGGAATAATTCACACTGGAATTGAC--- 114
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCCTTCATCCAGATCTCCAGGAAGTA----- 144
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGACATGA 204
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGTACTGAGCAGCAAGTAATGGCAAGTACAGGA 264
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGTTCTAGGTCCGCGATGTTCTGGAAGC 324
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaThrAsnAlaGlyAla 122
Db 325 ATATCTACTATATAATTAAAGGAGTTAGTGGCGCGCTGTGATAACAAGATAGTACGGA 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 385 AAGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTCAGCGTCTAAGT 444
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGACGCGTGGGATGCTGATAGTTGTTGTTGCTGCCCTGGAACAGT 504
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTTAACAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGATTATTACAGTTGA 564
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 565 GCGGTTGACAAAGTAT-----GATGTTATAACA 591
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGCGCAACTGCGAGCGGCGCTTAAGCTGAGGTGTTGTTCTCCA 651
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 242
Db 652 GGAACTGGATATTGTCGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 702
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCTCTCAGTAGTGGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTGCAGCCCTCTGCTCCAA-----GCACCCCGAGCTGGACTCCAGACAAAGTA 813
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAGCCAGATGAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGGTAGGTTAATGCATACAGGCTATAAAC----- 912
```

```

QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
Db 913 TACGATAACTATGCAAGCTAGTGTCTACTGATATGTCACCAAGAGGCGACCAACT 972
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrAspAla 353
Db 973 CACCAGTTCGTTATTAGCGAGCTTCGTTGCAACTGCCAATTATCTGGGCAATGCC 1032
QY 354 ProGlySerThrThrAlaSerThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1033 AAT-----ACGACCTTGATCTTACTCTACGAT 1062
QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnThrAsp 393
Db 1063 CCCAATGGAACACAGCTT-----GACTACTCTTAC-----ACC 1095
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db 1096 GCTACTATGGATTCGAAAGGTTGTTTATTACCAACCCCACTGATGGACATGGACAATT 1155
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1156 AAGGTTGTAAGCTAC-----ACCGAAGTGCACAACTATCAAGTAGATGTGTA 1203

```

RESULT 8

```

US-08-894-818B-4
; Sequence 4, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
;
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.
; US-08-894-818B-4
;
; Alignment Scores:
; Pred. No.: 11e-32 Length: 1566
; Score: 422.50 Matches: 138
; Percent Similarity: 43.96% Conservative: 55
; Best Local Similarity: 31.44% Mismatches: 167
; Query Match: 18.69% Indels: 79
; Ds: 3 Gaps: 17
;
; US-09-985-689A-5 (1-433) x US-08-894-818B-4 (1-1566)
;
; QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
; Db 58 AACTTGGGATATGATGTTCTCGAATCAACAATAGGAATAATTGACACTTGAATTGAC--- 114
;
; QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
; Db 115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA----- 144
;
; QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
; Db 145 ATTGGGTGGGTAGATTTTGTCAATGTTAGGAGTTATCCATACGATGACCATGACATGGA 204
;
; QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
; Db 205 ACTCATGTAGCTTCAATAGCAGCTGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 264
;
; QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
; Db 265 ATGGCTCAGGAGTAAAGCTGGCGGGAATTAAGGTTCTAGGTCCGATGTTCTTGGAGC 324
;
; QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 122
; Db 325 ATATCTACTATATTAAAGGAGTTAGTGGCGCTTGATACCAAGATAAGTACGGAATT 384
;
; QY 123 ArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
; Db 385 AAGGTCATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT 444
;
; QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
; Db 445 CAGGCTGTTAATGACGCGTGGGATGCTGGATTAGTTGTTGTTGCTGCGCTGGAACAGT 504
;
; QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
; Db 505 GGACCTAACAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTAATATTACAGTTGGA 564
;
; QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
; Db 565 GCGGTTGCAAGTAT-----GAACTAGCATGGGTCAACCA 591
;
; QY 203 GlnPheSerSerArgGlyValAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
; Db 592 AGCTTCTCAGCAGAGGCGCAACTCCAGCAGCGAGCTTAAGCTAGGTTGTTGCTTCCA 651
;
; QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 242
; Db 652 GGAAGTGGATAAATGCTGCCAGAGCAAGT-----GGAAGTGGATGGGTCAACCA 702
;
; QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
; Db 703 ATTAATGACTATTACACAGAGCTCCTGGGACATCAATGGCACTCTCTCAGTAGTGGT 762
;
; QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
; Db 763 ATTGAGCCCTCTTGTCTCAA-----GCACACCGAGCTGGACTCCAGACAAAGTA 813
;
; QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298

```

	814	AAACAGCCCTCATTAGAAGCAACTGCTATATCGTTAAAGCCAGATGAATAATAGCCGATATAGCC	873
Db	::::: :::::	::::: :::::	::::: :::::
	299	TyrProSerGlyAspGlnGlyTpdGlyAtgValThrLeuAspLysSerLeuAsnValala	318
Qy	::::: :::::	::::: :::::	::::: :::::
	874	TAC-----GGTGcaggtTAGGGTTAATGCATACAAGGCTATAAAC-----	912
Db	::::: :::::	::::: :::::	::::: :::::
	319	TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----	Thr 333
Qy	::::: :::::	::::: :::::	::::: :::::
	913	TACGATACTATGCAAAAGCTAGTGTTCACGTGGATATGTTGCCAACAAAGCGCAGCCAACCT	972
Db	::::: :::::	::::: :::::	::::: :::::
	334	TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla	353
Qy	::::: :::::	::::: :::::	::::: :::::
	973	CACCAGTTTCGTTATTAGCGAGCTTCGTCGTAACTGCCCACTATTACTGGGACAATGCC	1032
Db	::::: :::::	::::: :::::	::::: :::::
	354	ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla	373
Qy	::::: :::::	::::: :::::	::::: :::::
	1033	AAT-----AGCGACCTTGACTCTTACTCTACGAT	1062
Db	::::: :::::	::::: :::::	::::: :::::
	374	ProAsnGlyGlnLysTyrValGlyAsnAspPheSeryTrpTyrAspAsnAsnTrpasp	393
Qy	::::: :::::	::::: :::::	::::: :::::
	1063	CCCAATGGAAACCAGGTT-----GACTACTCTTAC-----	ACC 1095
Db	::::: :::::	::::: :::::	::::: :::::
	394	GlyArgAsnAsnValGluAsnValPheIleAsnAlaproGlnSerGlyThrTyThrIle	413
Qy	::::: :::::	::::: :::::	::::: :::::
	1096	GCTACTATGATTGCGAAAGGTGGTTATTACACCCCACTGATGGACATGGACAAATT	1155
Db	::::: :::::	::::: :::::	::::: :::::
	414	GluValGlnAlaTyrrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal	432
Qy	::::: :::::	::::: :::::	::::: :::::
	1156	AAGGTTTTGAAGCTAC-----AGCGAAGTGCAAACTATTCAAGTAGATGCGTA	1203
Db	::::: :::::	::::: :::::	::::: :::::

RESULT 9

US-08-894-818B-34
; Sequence 34, Application US/08894818B

; PATENT NO. 6261822
; GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MITTA, Masanori

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNASAWA, Susumu

APPLICANT: KATO, Ikunoshin

APPLICANT: KATO, IKUNOSHIN
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

; TITLE OF INVENTION: HYP
 : NUMBER OF SEQUENCES: 42

NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Str

CITY: Washi

STATE: D.C.

STATE: D.C.
COUNTRY: UR

COUNTRY: United States
ZIP: 20004

ZIP: 20004
COMPUTER READABLE FORM:

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk

```

```

; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible

```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MC-DOS
```

OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, v

```

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: U

FILING DATE: 20-MAY-1987

CLASSIFICATION: 435

CLASSIFICATION: 433
PRIOR APPLICATION DATA:

;; PRIOR APPLICATION DATA: 1
: APPLICATION NUMBER: 1

APPLICATION NUMBER: PCT/JF96/03255
FILING DATE: 07-NOV-1996

FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA:
 APPLICATION NUMBER: ID 2

APPLICATION NUMBER: JP 37

FILING DATE: 12-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,61

Db	1099	ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCACGTAGTGGT	1158
Ov	263	AsnValAlaGlnLeuArgGluHisSphelleLysAsnArgGlyIleThrPro-----	279

Alignment Scores:

Alignment scores:	
Pred. No.:	1.55e-32
Length:	1962

Pred. No.:	1.55e-32	Length:	1962
Score:	422.50	Matches:	138

Score:	422.50	138
Percent Similarity:	43.96%	55
Conservative:		
Matches:		

Percent Similarity:	43.96%	Conservative:	55
Best Local Similarity:	31.44%	Mismatches:	167

Best Local Similarity:	31.44%	Mismatches:	167
Query Match:	19.69%	Indels:	79

Query Match:

US-09-985-689A-5 (1-433) x US-08-894-818B-34 (1-1962)

QY	16	AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr	35
Db	454	AACTTGGGATATGATGTTCTGGATCAACAATAGGAATAATTGACACTGGAATTGAC---	510
QY	36	GlyArgAsnAspSerSerMetHisGluAlaPheAaGGlyLysIleThrAlaLeuTyrAla	55
Db	511	---GTTCTTCATCCAGATCTCCAAGGAAGAATA-----	540
QY	56	LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly	69
Db	541	ATTGGGTGGGTAGATTGTTGCAATGTTAGGAGTTATCCATACGATGACCATGGACATGGA	600
QY	70	ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly	84
Db	601	ACTCATGTAGCTTCAATAGCAGCTGTGTACTGGAGCAGCAAGTAATGGCACTACAGGGA	660
QY	85	MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AppSerSerGlyGly	102
Db	661	ATGGCTCCAGAGTAAGCTAGCTGGCGGAAATTAAGGTTCTAGGTGCGGATGGTCTCTGGAAGC	720
QY	103	LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla	122
Db	721	ATACTACTATATAATTAGGAGTTGAGTGGCGCTTGATACAAAGATAAGTACGGAATT	780
QY	123	ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg	142
Db	781	AAGGTCATTAATCTTCTCTCTGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT	840
QY	143	GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu	162
Db	841	CAGGCTGTTAATCAGCGCTGGGATGCTGATTAGTTGTTGTTGTCGCGCTGGAAACAGT	900
QY	163	GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly	182
Db	901	GGACCTAACAGTATACAATTCGGTTCTCCAGCAGCTGCCAAGCAAGTTATTACAGTTTGGGA	960
QY	183	AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla	202
Db	961	CCCGTTGACAAGTAT-----GATGTTATAACA	987
QY	203	GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro	222
Db	988	AGCTTCTCAAGCAGAGGGCCCACTGCGACGCGAGCTTAAGCCCTGAGGTTGTGCTGCCCA	1047
QY	223	GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn	242
Db	1048	GGAACTGGATTAATGTGTCGACAGCAAGT-----GGAACTAGCATGGGTCAACCA	1098
QY	243	TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly	262
Db	1099	ATTATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCCCTCAGCTAGCTGGT	1158
QY	263	AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro-----	279

QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1459 CCAATGGAAACAGGTT-----GACTACTCTTAC-----ACC 1491
QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrile 413
Db 1492 GCTACTATGGATTCCAAAGGTTGGTTATTACACCCCACTGATGGAAACATGGCAATT 1551
QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTTGTAAAGTAC-----ACCGGAAGTGCACAACTATCAAGTAGATGGGTA 1599

RESULT 11

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyocho
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-894-818B-6

Alignment Scores:
Pred. No.: 1.99e-30
Score: 401.50
Percent Similarity: 43.96%
Best Local Similarity: 29.61%
Query Match: 17.76%
DB: 3
Length: 1977
Matches: 130
Conservative: 63
Mismatch: 167
Indels: 79
Gaps: 17

US-09-985-689A-5 (1-433) x US-08-894-818B-6 (1-1977)

QY 16 AsnTyrGlyLeuTyrGlyGlnValValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTTGGGATATGATGTTCTTGAATCAATAGGAATAATGACACTGGAATTGAC--- 510
QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTTCTCATCCAGATCTCCAGAAAGTA----- 540
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGTGGGTAGATTTTGTCAATGGTAGGATTATCCATACGATGACATGACATGGA 600
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAG 660
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGTTCTTGAAGC 720
QY 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 122
Db 721 ATATCTACTATAATTAAGGGAGTTGAGTGGCCCTTGATACAAAGATAAGTACGGAATT 780
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 781 AAGTCTAATTAATCTTCTTGTTCGAGCAGAGCTCCGAGGAGCCGACCTCCCTCAGT 840
QY 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCGGTCAACAACGCTGGGACGCGGTATAGTCTGCTGCGCGCGCGGCAACAGC 900
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGGCGGAACACTACACGCTGGCTCACCCTCCCGGAGCAGGTCATACACCTCGGT 960
QY 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 961 GCA-----GTTGACACGACCAACATCGCC 987
QY 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAGCGGGGACCCGCGGACGAGAGGTCAAGCGGAGTGTGTCGCCCCC 1047
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsn 242
Db 1048 GCGGTTGACATCATAGCCCGCGCGCCAGC-----GGAACCCAGCATGGGCACCCCG 1098
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATAAACGACTACTACCAAGGCTCTGGAACACGATGGCCACCCCGCACGTTTCGGCC 1158
QY 263 AsnValAlaGlnLeuArgGluPheIleLysAsnArgGlyIleThrPro----- 279
Db 1159 GTTGGCGGCTCATCTCCAG-----GCCACCCGAGCTGGACCCCGCAAGGTG 1209
QY 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 1210 AAGACCGCCTCATCGACCGCGACATATGCGCCCCCAAGGAGATAGCGACATCGCC 1269
QY 299 TyrProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 1270 TAC-----GGTCCGGTAGGTGAACGCTACAGGCC-----ATCAAG 1308
QY 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----LysAlaThr 333
Db 1309 TACGACGACTACGCCAAGCTCACCTTCCCGCTCCGTCGCGCAAGAGGAGCGCCACC 1368
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrThrAspAla 353
Db 1369 CACACCTTCGACGTCAGCGGGCGCCACCTTCTGTGCGCGCCACCTCTACTTGGAC----- 1422

QY 354 ProGlySerThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
 DB 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGAC 1458
 QY 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnTrpAsp 393
 DB 1459 CCAAC-----GGGAACGAGGTGTACTACTCTCACCGCTTACAC--- 1500
 QY 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
 DB 1501 -----GGCTTCGAGAGTTCGCTACTACACCCGACCGCGGAACCTGGACGTC 1551
 QY 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
 DB 1552 AAGGTCCTCAGCTAC-----AAGGGCGCGCGGAACACTACACGAGTCGACGTCGTC 1599

RESULT 12
 US-09-000-016-3
 ; Sequence 3, Application US/09000016
 ; Patent No. 6143541
 ; GENERAL INFORMATION:
 ; APPLICANT: Akira ARISAWA et al.
 ; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
 ; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
 ; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 ; STREET: 2033 K Street, N.W., #800
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/000,016
 ; FILING DATE: January 30, 1998
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-721-8200
 ; TELEFAX: 202-721-8250
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2539 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptomyces viridosporus
 ; STRAIN: A-914
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 338..2539
 ; IDENTIFICATION METHOD: E
 US-09-000-016-3

Alignment Scores:
 Pred. No.: 3 7e-25 Length: 2539
 Score: 350.50 Matches: 134

Percent Similarity:	41.29%	Conservative:	53
Best Local Similarity:	29.59%	Mismatches:	168
Query Match:	15.50%	Indels:	98
DB:	3	Gaps:	19

US-09-985-689A-5 (1-433) x US-09-000-016-3 (1-2539)

QY	18	GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAAspThrGlyLeuAspThrGlyArg	37
Db	998	GGCTACGACGGCAAGGGCGTGAGATGCTCCGCTCTGGACACCGGTGTGACACGAGC---	1054
QY	38	AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly	57
Db	1055	-----CATCCGACCTGAAGGGCGGGTACCGCGTCCAAAGAACTTCACC	1099
QY	58	ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu	77
Db	1100	GGCGCCCGCGCGCGCGGACAAAGTGGCCACCGGACCCACGTCGCTCGATCGCGCG	1159
QY	78	GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal	92
Db	1160	GGCACGGGCGCCAGTCACAGGGCAAGTACAGGGCGTGCACCCGGCGCGCGATCTCTC	1219
QY	93	PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr	112
Db	1220	AAACGGCAAGTCTCTCAGCACTCC-----GGTTCCGGCGACACTCCGGCATC	1267
QY	113	LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly	130
Db	1268	CTCGCGGATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGTCTCACCATGAGCTGGGC	1327
QY	131	Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln	143
Db	1328	GGCATGGACACACCGGAGACCGACCGCGTGGAGGGCGCG-----	1366
QY	144	ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly	163
Db	1367	GTGCACAAAGTGTCCCGCGGAGGGCGTCTGTTCCGCATCGCGCGCGCAACGAGGCG	1426
QY	164	ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla	183
Db	1427	CCGAGTCTG-----ATCGGTTCGCGCGGAGCGCGGAGCGCGCGCTCCCGTCCGCGCGC	1480
QY	184	ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln	203
Db	1481	-----GTGCACGACAAAGGACAAAGTCTCGCGAC	1507
QY	204	PheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAlaPro	222
Db	1508	TTCTCTCCACCGGCGCCCGCTCGGCGACGGCGCCATCAAGCGGACGCTCACCGTCCC	1567
QY	223	GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn	242
Db	1568	GGCGTGGACATCACGCGCGCTCGGCGGGGCAACGACATCGCGCAGGAGTGGGTGAG	1627
QY	243	TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly	262
Db	1628	GGACGGCGCGCTACATGACCATCTCCGGCACGTCGATGGCGCACCCCGCACGTCGCGGC	1687
QY	263	AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer	282
Db	1688	CGCGCGCGCTCTCTGAAGCAGCAG-----CACCCCGAC	1720
QY	283</		

Db 1838 ATCCGCGACCGGCTCGGTGAGCTTCGGCTCCAGAGTGGCGGCACACCGACGAG 1897
Qy 330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
Db 1898 CCGGTACACAGCAGCTACCTACCGCACTCCGCACTCCGCGGCGGCGGCTTTCACGCTGAGGCTG 1957
Qy 347 SerLeuValTrrThrAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
Db 1958 AGCTGACCGCCACCGACCCCAAGGGCAAGGGCGGCGGCGGCTTTCACGCTGAGGCT 2017
Qy 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
Db 2018 -----GCCACACCGGTGACCGTCCGCGGCGGCGGCG----- 2047
Qy 385 SerTyrProTyrAspAsnAsnTrrPaspGlyArg-----AsnAsnValGlu 399
Db 2048 AGCGCTCGGTGACATGACCGCCGACACCGCGGCGGCGGCGGCGGCGGCGGTAC 2107
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
Db 2108 TCGGCGTACGTGCTGCC---ACGGCGCGGCGGCGAGCGGTCCGCGCGCGCGGCTG 2164
Qy 414 -----GluValGlnAlaTrrAsnVal-ProSerGly 423
Db 2165 CAGCGCGAGGTGAGTACGACCGTACCGTCCGCG 2201

RESULT 13
US-09-514-340-3
Sequence 3, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914

FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3
Alignment Scores:
Pred. No.: 3-7e-25 Length: 2539
Score: 350.50 Matches: 134
Percent Similarity: 41.28% Conservative: 53
Best Local Similarity: 29.58% Mismatches: 166
Query Match: 15.50% Indels: 98
DB: 19
US-09-985-689A-5 (1-433) x US-09-514-340-3 (1-2539)
Qy 18 GlyLeuTyrGlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCGCAAGGGCGGTGAAGATCGCGCTCTCGACACCGGTGTGCACACGAGC--- 1054
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCCGACCTGAAGGGCGGGTGACCGGTCCTCAAGAACTTCACC 1099
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGGCGGCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGG 1159
Qy 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1219
Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAAGGTCTCTCGACGACTCC-----GGTTTCGGCGACGACTCCGCGCATC 1267
Qy 113 LeuPheSerGlnAlaTrrAsnAla-----GlyValaArgIleHisThrAsnSerTrrGly 130
Db 1268 CTCGCGGCGATGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1327
Qy 131 Ala-----ProValAsnGlyValaTrrThrAlaAsnSerArgGln 143
Db 1328 GGCATGGACACACCGGAGACCGACCGCTCGAGGCGGCG----- 1366
Qy 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCACAAAGCTGTCCGCGCGGAGAGGGCGCTCTGTTCGCCATCGCGCGCGGCGGCGGCGG 1426
Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCTG-----ATCGGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1480
Qy 184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
Db 1481 -----GTCGACGACAGGACAGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1507
Qy 204 PheSerSerArgGly-----AlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1567
Qy 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrrAlaAsn 242
Db 1568 GCGGTGGACATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1627
Qy 243 TyrAsnSerLysTrrAlaTrrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1688 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1720


```

QY 204 PheSerSerArgGly---AlaThrArgAspGlyAArgIleLysProAspValThrAlaPro 222
Db 1508 TTCTCTCCACCGCCCGCCCGCTCGCGACGGCCATCAAGCCGCGCTCACCGCTCCC 1567
QY 223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGTGGACATACCGCCCGCTCGCGGAGGACGACATCGCGCAGGAGTGGTGAG 1627
QY 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACGGCGCGGTACATCACCATCTCCGCGACGTGATGGGACCCCGACGTGCGGGC 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
Db 1698 GCGCGCGCCCTCTGAGCAGCAG-----CACCCCGAC 1720
QY 283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TCGACCTCCGCGAACTCAAGCGCGCTCACCGCTCCACCAAGGGCGGC---AAGTAC 1777
QY 300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGTTCGACGAGGCTCGCGCGGATCCAGCGGACAGGCGCTCCAGCAGACCGTG 1837
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
Db 1838 ATCGCGACCGCTCTCGGTGAGTTCGCGCTCCAGCAGTGGCGCACACGACGACGAG 1897
QY 330 -----GlnIysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuIle 346
Db 1898 CCGGTACCAAGCAGCTGACCTACCGACACCTCGGACCCAGCAGCGTCACGCTGAAGTG 1957
QY 347 SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
Db 1958 ACGTCACCGCACCGACCCCAAGGCGGCGCGCGGCTTCTTACGCTGGGC 2017
QY 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
Db 2018 -----GCCACCAAGTGTACCGTCCCGCGGCGCGC----- 2047
QY 385 SerTyrProTyrAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
Db 2048 AGCGCTCGTGCATGACGATGCCCGACCGCGCGCGCGCGCGCGCGCGCGTAC 2107
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
Db 2108 TCGCGTACGTGTGCGC---ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2164
QY 414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
Db 2165 CAGCGCGAGGTGAGTGTGACGACGTGACCGTCCCGCGCGCGC 2201

RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Wandoroth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS

```

```

SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 1,37e-24 Length: 2809
Score: 345.50 Matches: 133
Percent Similarity: 41.06% Conservative: 53
Best Local Similarity: 29.36% Mismatches: 169
Query Match: 15.28% Indels: 98
DB: 4 Gaps: 19

US-09-985-689A-5 (1-433) x US-09-514-340-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCGCAAGGCGGTGAAGATCGCGTCTCGACACCGGTGTGACACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGCGGCGGTGACCGGTCCCAAGAACTTCACC 1099
QY 58 ArgThrAsnAlaAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1219
QY 93 PheGlnSerIleMetAspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAGGTCTCTGACGACTCC-----GGTTTCGCGCGCGACTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTCGCGCGCATGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
QY 131 Ala-----ProValAsnGlyValaTyrThrAlaAsnSerArgGln 143

```

1328 GGCATGACACACCGGAGACCGACCGCTGGAGGGGGC----- 1366
144 ValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
1367 GTCGACAAAGCTGCCCGAGAGAGGGCTCTGTCGCCATCGCGCGGCAACGAGGGC 1426
164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
1427 CCGGAGTCG-----ATCGGTTCGCGCGGAGCGCGGCGCCCTCACCGTCGCGGCC 1480
184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
1481 -----GTCGACGACAAAGCAAGCTCGCGGAC 1507
204 PheSerSerArgGly---AlaThrArgAspGlyArgIleIysProAspValThrAlaPro 222
1508 TTCTCTCCACCGCGCCCGCTCGCGGACGCGGCCCATCAAGCCGACGTCACCGCTCCC 1567
223 GlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
1568 GCGGTGACATACCGCGCGCTCGCGGAGGGCAACGACATCGCGCGGAGGTCGGTGAG 1627
243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
1628 GGACCGCGCGCTACATGACATCTCCGGCACGTCGATGCGGACCCCGCACGTCGCGGGC 1687
263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSer 282
1688 GCGCGCGCTCTCTGAAGACGAG-----CACCCCGAC 1720
283 LeuIleLysAlaAlaLeu-----IleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
1721 TGGACCTCCGCGCAACTGAGGGCGCGCTCACCGGCTCCACCAAGGCGGC---AAGTAC 1777
300 ProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
1778 ACCCGTTCGAGCAGGGTTCGGCGCGATCCAGGCCACAGGCGCTCCAGCAGACCGTG 1837
320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
1838 ATCGCCGACCGGTCTCGGTGAGCTTCGGCTCCAGCAGTGGCGGCACACCGACGACGAG 1897
330 -----GlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
1898 CCGGTACCAAGCAGCTACCTACCGCACTCGGACCCCGACCCAGGACGTCACGCTGAAGCTG 1957
347 SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
1958 ACGTCGACCGCCACCGACCCCAAGGCAAGCGCGCGCGGCTTCTTCAAGCTGGGC 2017
365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
2018 -----GCCACACCGGTACCGTCCCGCGGGGGC----- 2047
385 SerTyrProTyrAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
2048 AGCGGCTCGTCCGATACCGCCGACACCGGCTCGCGGCGCACGCTGGAGCGGCGGTAC 2107
400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle----- 413
2108 TCGGCGTACGTGCTGCC---ACGGCGCGGCGGACGCTCCGCGACGCGCGCGCGGTG 2164
414 -----GluValGlnAlaTyrAsnVal-ProSerGly 423
2165 CAGCGCGAGTCTGAGTCTGACGCTGACCGTCCCGC 2201

Search completed: April 4, 2004, 12:02:59
Job time : 92.4512 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-5

Perfect score: 2261

Sequence: 1 NDVARGIVKADVAQNNGY.....EVQANVPSGQRFSLAIVH 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_31032004_161809_4271/app_query.fasta_1.3498
-DB=Published Applications NA -QWTF=fastap -SUPFLX=rnpb -MINMATCH=0.1
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0985689.OCGN_1_1_601.0runat_31032004_161809_4271
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	1998.5	88.4	1305	15	US-10-385-662-1	Sequence 1, Appli
2	443	19.6	1977	13	US-10-090-624-11	Sequence 11, Appl
3	422.5	18.7	1236	13	US-10-090-624-2	Sequence 2, Appli
4	422.5	18.7	1962	13	US-10-090-624-15	Sequence 15, Appl
5	341.5	15.1	135638	14	US-10-314-657-1	Sequence 1, Appli
6	334	14.8	1329	9	US-09-974-300-1934	Sequence 1934, Ap
7	329	14.6	3417	14	US-10-156-761-3306	Sequence 3306, Ap
8	329	14.6	9025608	14	US-10-156-761-1	Sequence 1, Appli
9	327	14.5	3624	14	US-10-156-761-5701	Sequence 5701, Ap
10	327	14.5	9025608	14	US-10-156-761-1	Sequence 1, Appli
11	307	13.6	4785	13	US-10-090-624-5	Sequence 5, Appli
12	298.5	13.2	3743	10	US-09-927-827-29	Sequence 29, Appl
13	295	13.0	1560	15	US-10-084-846A-113	Sequence 113, App
14	295	13.0	59816	15	US-10-084-846A-1	Sequence 1, Appli
15	295	13.0	59816	15	US-10-084-846A-2	Sequence 2, Appli
16	284.5	12.6	3303	14	US-10-156-761-5384	Sequence 5384, Ap
17	276.5	12.2	2166	12	US-10-344-231-17	Sequence 17, Appl
18	276.5	12.2	2166	12	US-10-363-332A-17	Sequence 17, Appl
19	273	12.1	3788	10	US-09-927-827-33	Sequence 33, Appl
20	269	11.9	1140	8	US-08-322-678-11	Sequence 11, Appl
21	269	11.9	1140	16	US-10-323-324-11	Sequence 11, Appl
22	269	11.9	1143	14	US-10-313-853-6	Sequence 6, Appli
23	269	11.9	2588	12	US-10-344-231-20	Sequence 20, Appl
24	269	11.9	2588	12	US-10-363-332A-20	Sequence 20, Appl
25	268	11.9	1140	8	US-08-322-678-12	Sequence 12, Appl
26	268	11.9	1140	16	US-10-323-324-12	Sequence 12, Appl
27	260.5	11.5	1306	9	US-09-966-921A-1	Sequence 1, Appli
28	260.5	11.5	1330	9	US-09-966-921A-5	Sequence 5, Appli
29	259.5	11.5	840	14	US-10-209-812-1	Sequence 1, Appli
30	258.5	11.4	2267	12	US-10-344-231-18	Sequence 18, Appl
31	258.5	11.4	2267	12	US-10-363-332A-18	Sequence 18, Appl
32	252	11.1	1485	9	US-09-974-300-1938	Sequence 1938, Ap
33	251	11.1	3884	10	US-09-927-827-33	Sequence 34, Appl
34	250.5	11.1	1971	9	US-09-974-300-1935	Sequence 1935, Ap
35	247.5	10.9	2192	12	US-10-424-599-112429	Sequence 112429,
36	242	10.7	1140	15	US-10-146-905A-9	Sequence 9, Appli
37	240.5	10.6	3452	10	US-09-927-827-30	Sequence 30, Appl
38	238	10.5	1140	9	US-09-920-118-13	Sequence 13, Appl
39	231.5	10.2	4338	9	US-09-891-711-3	Sequence 3, Appli
40	230	10.2	4338	12	US-10-342-887-953	Sequence 953, App
41	230	10.2	1332	14	US-10-156-761-5689	Sequence 5689, Ap
42	230	10.2	1575	12	US-10-344-231-19	Sequence 19, Appl
43	230	10.2	1575	12	US-10-363-332A-19	Sequence 19, Appl
44	228	10.1	4198	9	US-09-891-711-5	Sequence 5, Appli
45	227.5	10.1	522	10	US-09-824-893A-98	Sequence 98, Appl

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSURISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

```

; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 6,99e-210 Length: 1305
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservative: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.39% Indels: 1
DB: 15 Gaps: 1

US-09-985-689A-5 (1-433) x US-10-385-662-1 (1-1305)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyr 20
Db 1 AATGATCTTGGCGTGGCAATTGTCAAAGCGGATGTGGCTCAGAGCAGCTACGGGTGTAT 60
Qy 21 GlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAAGACAGATCGTAGCGGTGCGGATACAGGCTTGATACAGTCCAAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAAGCTTCGCGGGAATTTACTGCAATATATGCAATTTGGGACGCAAT 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGAATGGTCATGTCAGCATGTCATGTCATGTCATGTCATGTCAT 240
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATAAGAAATGGCGCTCAGCGCAATCTAGTCCTCCAAATCATCATGATAGC 300
Qy 100 SerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsn 119
Db 301 GGTGGGGACTTGGAGGACTACCTTCAATTCGAAATCGAAACCTTATTCAGCCCAAGCATACAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAla 139
Db 361 CTGTGTGCCAATTCATCAAACTCTGGGAGCAGCATGATGGGCTTACACAACA 420
Qy 140 AsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 421 GATTCAGAAATGTGGATGACTATGTGGCGCAAAATGATATGACGATCCTTTTCGCTGCC 480
Qy 160 GlyAsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAlaIle 179
Db 481 GGGAAATGAAGACCAACCGCGGACCATAGTCAGCCAGCAGCAGCATTAATAATGCATA 540
Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 541 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGCGGACAAATATCAAC 600
Qy 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTGGCAGACTTCTCTACGTGGACCGCAAGGATGACGCGATCAAAACCGGATGTC 660
Qy 220 ThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCACCGGACGTTTATATCATCATCAGCAAGATCTTCTCTGACCGGATTCCTCTTC 720
Qy 240 TrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259

```

```

Db 721 TGGCGNACCATGACAGTAATAATGATCATCATGGTGGAGCGTCCATGGCTACACCGATC 780
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 279
Db 781 GTTGTCTGAAACCGTGGCAGCAGCTTCGTGAGCATTTTGTGAAACACAGAGGACATCACACCA 840
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 841 AAGCTTCTCTATTAAAGCGGCACTGATGCGGTCGAGCTGACATCGGCTTGGCTAC 900
Qy 300 ProSerGlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 319
Db 901 CCGAACGGTAACCAAGGATGGGACGAGTGCATTTGGATTAATCCCTGAAACGTTGCCAT 960
Qy 320 ValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGln 339
Db 961 GTGAACGAGTCCAGTCTCTATCCACCAGCCAAAGCGAGTACTGCTTTACTGCTACT 1020
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTyrThrAspAlaProGlySerThrThrAla 359
Db 1021 GCCGCAAGCCTTTGAAATCTCCCTGTATGCTGATGCTCCCTGCGAGCAACATGCT 1080
Qy 360 SerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyr 379
Db 1081 TCCGTAAACGCTTGCAATGATCTGACCTTGTCTATTACCGCTCCAAATGGCACACAGTAT 1140
Qy 380 ValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTAGAAATGACTTTACTTCCCATACATGATTAATCTGGATGGCGCAATTAAGTAGAA 1200
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 419
Db 1201 AATGATTTTAAATGACCAACAAAGCGGACGATTAACAATTTAGGTACAGCTTTAAAC 1260
Qy 420 ValProSerGlyProGlnArgPheSerLeuAlaIleValHis 433
Db 1261 GTACCGGTGGACACAGACCTTCTCGTTGGCAATTTGTAAT 1302

```

```

RESULT 2
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication NO. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozi
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 1.93e-38 Length: 1377
Score: 443.00 Matches: 137
Percent Similarity: 46.03% Conservative: 66
Best Local Similarity: 31.07% Mismatches: 170
Query Match: 19.59% Indels: 68
DB: 13 Gaps: 16

```

US-09-985-689A-5 (1-433) x US-10-090-624-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValValAla 27
DB 433 ATAGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGGAAGCGGTGTGTGCTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
DB 493 ATCTGCTGATACGGGTATAGACGGAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
DB 535 GCGAAGGTATAGGTGTGTAGACCGCTCAACGCGAGTCTCGTGGCGTCAAGGTCTCGGTGCGGATCTG 714
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn 82
DB 595 GGACACGGAACCCAGCTTGGCGGTATCTGTTCGCGAACCGGACGGAGGTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSer 100
DB 655 ATAGCGTCTGCGCGCGCGGCGAAGTCTCGTGGCGTCAAGGTCTCGGTGCGGATCTG 714
QY 101 GlyGlyLeuGlyGlyLeuProSerSerLeuAsnThrLeuPheSerGlnAlaThrPAsnAla 120
DB 715 GGAACGCTCCACCATCATCGCGGTGTGACTGGTCTCCAGAACCAAGCAAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
DB 775 GGGATAAGGTATCAACCTCTCCCTCGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
DB 835 CTCAGTCAGCGCGTCAACACACCGCTGGGACGCGGTATAGTCTGCGTCTGCGCGCGC 894
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 895 ACACGCGCGCAACACCTACACCTCGCTACCGCGCGCGCGGACGAGTATAC 954
QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
DB 955 GTGCGGTGCA-----GTTGACAGCAACGACCAAC 981
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
DB 982 ATCGCAGCTTCTCCAGCGGAGCGACCGCGGAGCGGAGGCTCAACCGGAGTCTGTC 1041
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
DB 1042 GCGCGCGGTGTGACATCATACCGCGCGCGCAGC-----GGAACACGATGCGC 1092
QY 241 AlaAsnTyrAsnSerLysThrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
DB 1093 ACCCGATTAACGACTACTACCAAGCGCTCTGGAACACGATGCGCACCGCGCACGTT 1152
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro--- 279
DB 1153 TCGGGGTTGGCGGTCTATCTCCAG-----GCCACCGCGAGTGGACCGCGGAC 1203
QY 280 -----LysProSerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGly--- 296
DB 1204 AAGGTGAAGACCGCCTCATCAGACCGCGCATAGTCCGCCCAAGGAGATACGCGAC 1263
QY 297 LeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAlaPheLysSerLeuAsn 316
DB 1264 ATCGCTAC-----GGTGGGTAGGTGAACGTCTACAGGCC----- 1302
QY 317 ValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGln-----Lys 331
DB 1303 ATCAAGTACGACGACTACGCAAGTCTACCTTCACCGGTCTGCTGCGCGCAAGGAAGC 1362
QY 332 AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThr 351

DB 1363 GCCACCCACACCTTCGACGTCAGCGCGCCACCTTCGTGACCGCCACCTCTACTTGGAC 1422
QY 352 AspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValle 371
DB 1423 -----ACGGGCTCGAGGACATCGACTCTACCTCTACCTC 1452
QY 372 ThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsn 391
DB 1453 TAGCAGCCCAAC-----GGGAACGAGGTGTGACTCTCTACACCGCCTAC 1497
QY 392 TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyr 411
DB 1498 TAC-----GGCTTCGAGAGAGTCTGCTACTACACCGCCACCGCGGAACCTGG 1545
QY 412 ThrIleGluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
DB 1546 ACGTCAAGGTCTGCTAGCTAC-----AAGGGCGCGGCACTACCGAGTCTGACGTC 1596
QY 432 Val 432
DB 1597 GTC 1599

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 1,76e-36 Length: 1236
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 13 Gaps: 17

US-09-985-689A-5 (1-433) x US-10-090-624-2 (1-1236)

QY 16 AsnTyrGlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
DB 58 AACTTGGGATATGATGTTCTGGAATCAACAATAGGAATATTCACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
DB 115 -----GTTCTCATCCAGATCTCCAGGAAGAATA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
DB 145 ATTGGGTGGGTAGATTGTCATGTAGTAGGTATTCATCATGATGACCATGACATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84

Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAGGGA 264
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 265 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGTTCTAGGTGCGGATGTTCTGGAAGC 324
Qy 103 LeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 325 ATATCTACTATAATTAAGGAGTTGAGTGGCGCGTTGATAACAAGATAAGTACGGAAT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpAlaAsnSerArg 142
Db 385 AAGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
Qy 143 GlnValAspGluTrpValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGACGCGTGGGATGCTGGATTAGTTGTTGGTTCGCCGCTGGAACAGT 504
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAACAGTATACAAATCGTTCTCCAGCAGCTGCAAGCAAGTTATTACAGTTGA 564
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 565 GCGGTTGACAAGTAT-----GATGTTATAACA 591
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyValGlyLeuLysProAspValThrAlaPro 222
Db 592 AGTTCTCAAGCAGAGGCGCAATGCGAGCGGAGGCTTAAGCTGAGGTGTTGTTGCTCA 651
Qy 223 GlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 GGAACTGGATAATGTCGACAGCAAGT-----GGAACTAGCAGTGGTCAACCA 702
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCAGTAGTGTGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 763 ATTCAGCCCTCTTGCTCCAA-----GCACACCGGAGCTGGACTCCAGACAAAGTA 813
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 814 AAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCC 873
Qy 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 874 TAC-----GGTGCAGGTAGGTTAATGCATACAGGCTATAAAC----- 912
Qy 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
Db 913 TACGATAACTATGCAAAAGCTAGTGTTCACCTGGATATGTTCCCAACAAAGCGCAACCACT 972
Qy 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAla 353
Db 973 CACCAGTTCGTATTAGCGGAGCTCGTTCGTAACCTGCCATATATAGTGGCAATGCC 1032
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuValIleThrAla 373
Db 1033 AAT-----AGCGACCTTGATCTTTTACCTCTACGAT 1062
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAsp 393
Db 1063 CCCAATGGAACCAAGTT-----GACTACTTTTAC-----ACC 1095
Qy 394 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIle 413
Db 1096 GCCTACTATGATTGCAAAAGGTTGTTATTACAAACCACTGATGGAACATGGCAAT 1155
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432

Db 1156 AAGTTTGAAGCTAC-----AGCGAAGTGCAAACTATCAAGTAGATGTGGTA 1203
RESULT 4
US-10-030-624-15
; Sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15
Alignment Scores:
Pred. No.: 3,45e-36 Length: 1962
Score: 422.50 Matches: 138
Percent Similarity: 43.96% Conservative: 55
Best Local Similarity: 31.44% Mismatches: 167
Query Match: 18.69% Indels: 79
DB: 17 Gaps: 17
US-09-985-689A-5 (1-433) x US-10-090-624-15 (1-1962)
Qy 16 AsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTGGGATATGATGTTCTGGAATCAATAGGAATAATGACACTGGAATTGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTTCTCCAGATCTCCAGGAAGAATA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsnLysGly 84
Db 601 ACTCATGTAGTTCATATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAATGACAGGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGCGGGAATTAAGTTCTAGGTGCGGATGTTCTGGAAGC 720
Qy 103 LeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAla 122
Db 721 ATATCTACTATAATTAAGGAGTTGAGTGGCGCGTTGATAACAAGATAAGTACGGAAT 780
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArg 142
Db 781 AAGTCAATTAATCTTCTCTGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 840
Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCTGTTAATGACGCGTGGGATGCTGATAGTTGTTGTTGTCGCGCTGGAACAGT 900
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182

```
Db 901 GGACCTAACAGTATACATCGTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGA 960
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisLeAla 202
Db 961 GCGGTGACAGTAT-----GAGTTATATACA 987
Qy 203 GlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaPro 222
Db 988 AGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGCTTAAGCTTGAGGTTGTGCTCCA 1047
Qy 223 GlyThrPheIleLysSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1048 GGAACCTGGATTAATGCTCCAGCAAGT-----GGACTAGCATGGGTCAACCA 1098
Qy 243 TyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValalaGly 262
Db 1099 ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCTCTCAGCTAGCTGGT 1158
Qy 263 AsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro----- 279
Db 1159 ATGTCAGCCCTCTTGCTCCAA-----GCACACCCGAGCTGGAGCTCCAGACAAAGTA 1209
Qy 280 LysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGly 298
Db 1210 AAAACAGCCCTCATGAAACTGCTGATATCGTAAAGCCAGATGAATAAGCCGATATAGCC 1269
Qy 299 TyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 318
Db 1270 TAC-----GGTCAGGTAGGGTTAATGATACAGAGCTATAAAC----- 1308
Qy 319 TyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAla-----Thr 333
Db 1309 TACGATAACTATGCAAGCTAGTGTCTACTGATATGTTGCCAACAGGCGCAAACT 1368
Qy 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValThrTrpAspAla 353
Db 1369 CACCAGTTCTGTTATTAGCGGAGCTCGTTTCGTAACCTGCCATATTACTTGGGACAATGCC 1428
Qy 354 ProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 1429 AAT-----AGCGACTTGATCTTACTCTACCAT 1458
Qy 374 ProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnTrpAsp 393
Db 1459 CCCAATGGAACACAGTT-----GACTACTCTTAC-----ACC 1491
Qy 394 GlyArgAsnAsnValGlnAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIle 413
Db 1492 GCTACTATGGAATTCGAAAGGTGGTTATTACAAACCCCACTGATGGAACTGGCAATT 1551
Qy 414 GluValGlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1552 AAGGTTGTAAGCTAC-----AGCGAAGTGCAAACTCAATCAAGTAGATGTGGTA 1599
```

RESULT 5

```
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; FILE OF INVENTION: Syntheses and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1
```

```
Alignment Scores:
Pred. No.: 1,31e-24 Length: 135638
Score: 341.50 Matches: 139
Percent Similarity: 42.17% Conservative: 55
Best Local Similarity: 30.22% Mismatches: 155
Query Match: 15.10% Indels: 111
DB: 14 Gaps: 23
```

US-09-985-689A-5 (1-433) x US-10-314-657-1 (1-135638)

```
Qy 6 GlyIleValLysAlaAspValAlaGlnAsnAsn----- 16
Db 8067 GGCAGGGTGAAGCCGATCTGGCCGACTCCACGCCAGATCGGCGCGCAGAGGTATGG 8126
Qy 17 ---TyrGlyLeuTyrGlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGAGGGCCACACCCGCCAGGACGTGAAGTCCGATGCTCCAGCGCGCGGACACC 8186
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
Db 8187 -----GNACACCCGACCTGCTGGGCGAGGTCCGACAGCGCCAGC 8228
Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisIleThrHisValAla 73
Db 8229 TTGCTCCCGCGGAGGACGACATCGCC---GACTACACGCGCCAGCGACGACGTCGCC 8285
Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
Db 8286 TCGACCATCTGCGCAGCGGCAGCGCTCCGACGGCAGGAGCGGGTGTGCGCTCGGC 8345
Qy 89 AlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSer 108
Db 8346 GCGCGGTGTCTCGCAAGGTGCTCAACTCCAGGGC-----AGCGCCAGGAATCG 8399
Qy 109 AsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHis 125
Db 8400 TGGATC-----ATCGGGGCGCATGGAGTGGCGCCCGCCGACAGAGCCAGGATCATC 8453
Qy 126 ThrAsnSerTrpGlyAlaProValAsnGlyValatYrThrAlaAsn-----SerArg 142
Db 8454 AGCATGAGCCCTGGGC-----GGCGGCGGTGACAAAGAACACCGCATGAGCCAG 8501
Qy 143 GlnValAspGlnTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 8502 GCGGTCCAGCAACTCAGCCACGACGCGGCGCTTGTGCTGATCGCGCGGCAACGCG 8561
Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 8562 GGCCCGCACTCC-----ATCAGCAGCCCGGTCGCGCAGACTCCGCGTACCGTCGCG 8615
Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAla 202
Db 8616 GCC-----GTCCACTCCACCGACGCTCGCC 8642
Qy 203 GlnPheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 8643 GACTTCTCCAGCAGGCGCCCGCTGACGCGGACGCGGGGCTGAAGCCGGAGATCACCGCG 8702
Qy 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 8703 CCGCGCTCGACATCGTTCGGGGCGGCTCGCACTACAAGCGCGGCTCCGCGCTAC----- 8756
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 8757 -----TACACCACGATGAGCGGCGACGTCGATGGCGACCGCGCAGCTCGCC 8801
```

Qy 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyLeuThrProLysPro 281
|||
Db 8802 GGTGTCCGCGCTCTCGCGCCGAGCAGCCCGAGTGGACCGGACCCAGCTCAAGGAG 8861
|||
Qy 282 SerLeuIleLysAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSer 301
|||
Db 8862 GCACCTGGTCAGCAGCCCAAGCAACCGCGCG- -TACACCCCG 8903
|||
Qy 302 GlyAspGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsn 321
|||
Db 8904 TACCAGCGGCGCGCGCTCGACGCGCGCGCGCTGCACACACCGCTCTCGCC 8963
|||
Qy 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
|||
Db 8964 ACCACGACCGCTTACTCC- -GGCTTCCAC- - 8990
|||
Qy 342 LysProLeuLysIleSerLeuValThrAspAlaProGlySerThr- -Thr 358
|||
Db 8991 - - - - -ACGTGCCCCCAAGCCCGGGGAGACCGATCTCCGGACG 9029
|||
Qy 359 AlaSerTyrThrLeuValAsnAsp- - - - -LeuAspLeuValIle- -Thr 372
|||
Db 9030 GTACGCTACACCAACGTCGCGACGCGCGCTCAGCTCAACCTGCGCTCAACGCGCAC 9089
|||
Qy 373 AlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTyr 392
|||
Db 9090 GTCCGCGCGGGTG- - - - -TTCAGCTCTCCGAGACCAT- - - - - 9125
|||
Qy 393 AspGlyArgAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412
|||
Db 9126 - - - - -GTCACCGTCCCGCGCAGCGCACCGCGCACG 9155
|||
Qy 413 IleGluVal- - - - -GlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
|||
Db 9156 GTACCCCTGACCGCGCTCTGGCAAGCTGCGCGGCGACCGTCTGCGTCAAGCGCGTGATC 9215
|||

RESULT 6

US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085,500-US
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:

Pred. No.:	1,08e-26	Length:	1329
Score:	334.00	Matches:	105
Percent Similarity:	44.58%	Conservative:	43
Best Local Similarity:	31.63%	Mismatches:	116
Query Match:	14.77%	Indels:	68
DB:	9	Gaps:	12

US-09-985-689A-5 (1-433) x US-09-974-300-1934 (1-1329)

Qy 11 AspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnValValAlaValAlaAsp 30
|||
Db 406 GAAGTGGTCAAGAAACATCAGACGCTGACAGGCAAGGAGTGACAGTCTGCTGATGAT 465
|||

RESULT 7

US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN

Qy 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle 50
|||
Db 466 ACGGCGTA- - - - -TACCTTCAGGAAGATCTTGAAGGACGATC 504
|||
Qy 51 ThrAlaLeuTyr- - - - -AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGly 69
|||
Db 505 AGGCTTTTCAAGACTTTATCAACACGAGAGAACAGACCTTATGATCAATGGGACGCG 564
|||
Qy 70 ThrHisValAlaGlySerValLeuGlyAsnAlaLeu- - - - -AsnLysGly 84
|||
Db 565 ACACACTCGCGCGGTGATGCTTGGGAAACGGAGCGGCTCATCGGTGATGATCAATCGCGGA 624
|||
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
|||
Db 625 CCTGCTCTCAAGCAGAACTTGTGGGTAAAGTATTTGGACAAATGGATCC- -GGA 681
|||
Qy 105 GlyLeu- - - - -Pro 107
|||
Db 682 TCGTCTCAAAACCGTCATTCAAGCGGTAGATTGGTGCATTCAATTAAGGAAATCCT 741
|||
Qy 108 SerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAlaArgIleHisThrAsn 127
|||
Db 742 GATGATCGGATCGACATTTATTTCAATGTCTATTTGGGTGACAGAGCCTTGCCTACGAGAT 801
|||
Qy 128 SerTrpGlyAlaPro- - - - -ValAsnGlyValaTyrThrAlaAsnSerArgGln 143
|||
Db 802 GAAGAAGATCCAGTCGCTTAAAGCTGTTCATGCAGCATGGGACGACGCG- - - - - 852
|||
Qy 144 ValAspGluTyrValaArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
|||
Db 853 - - - - -ATTGTTGTATGTGCGCAGCGGCACTCCGCT 885
|||
Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
|||
Db 886 CCGTATGCGCAACAGATTCCAGCCCGGTGTGACGACGAGATTATTACATCGCGAGCC 945
|||
Qy 184 ThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIleAlaGln 203
|||
Db 946 TTGGATGACAGG- - - - -GATACAGTCAGCGGAGGATGACGATGCGCTCT 993
|||
Qy 204 PheSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
|||
Db 994 TATTCAAGCAGAGCGCGACAACTTATGTCAGTCAACCGGACTTGTCTGTACCGGCG 1053
|||
Qy 224 ThrPheIleLeuSerAlaArgSer- - - - -SerLeuAlaProAspSerSerPheTrpAla 241
|||
Db 1054 ACAAATATTACGTCGCTCGTTCACCGGATCTTTCTCGATAAGCTGCAAAAAACAAAC 1113
|||
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
|||
Db 1114 AGAGTCGGCACAATAATATGACATGTCCGGAACCTCGATGGCTACGCGATCTCGCA 1173
|||
Qy 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
|||
Db 1174 GGAATTCGCGCA- - - - -CTTATCTTCAGCAAGCCCGGCGCACAGAACCT 1218
|||
Qy 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspVal- - - - - 295
|||
Db 1219 GATGAAGTCAAAACAGCTGCTTAATGGACGCT- - - - -ACCGATTATGGAAGATCGCGATCCA 1275
|||
Qy 296 - - - - -GlyLeuGlyTyrProSerGlyAspGln 304
|||
Db 1276 AATGTTTACGTCAGCGGTATCATCAACGACGACAA 1311
|||

```

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 15e-25 Length: 3417
Score: 329.00 Matches: 114
Percent Similarity: 45.27% Conservative: 44
Best Local Similarity: 32.66% Mismatches: 126
Query Match: 14.55% Indels: 67
DB: 14 Gaps: 12

US-09-985-689A-5 (1-433) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
DB 745 GGGTACGACGGAAGGGCGTCGAAGATCGCGCTCGTGACACCGGTGTGCAC----- 795
QY 38 AsnAspSerMetMetGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 796 -----GCGACCCACCGGACCTCAGAGCAGGAGTGGCGGAGTCCAGAGACTTCTCC 846
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 847 GCGCGCGCGGCGCGCGCGGACCATCTCGGTCTACGCGCAGCAGCAGCTGCGTCCATCCGCGG 906
QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 907 GGACCGCGCGCGCAAGTCCACGCGCAGTACAGAGGTGTGCGCGCGCGCGCGGAGTCTC 966
QY 93 PheGlnSerIleMetAspSerSerGly-----GlyLeu 103
DB 967 AACGGCAAGTCTCTCAGCAGCAGCAGCGGCTCGGCGCAGCAGTCTCGCGCGCATG 1026
QY 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAlaGlyAla 123
DB 1027 GAGTGGCGCGCGCAGCAGCGCGCGCGCGCTCGTCACTGAGCTGGCGCGCGCGCAGCACC 1086
QY 123 gileHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAlaAsnSerArg 143
DB 1087 CCGGAGATCGACCGCGTGGAA-GCGGAGGTCAACAGCTC----- 1125
QY 143 nValAspGluTyrValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
DB 1126 -----TCCAGGAGAGAGGCGATCTCTTTCGCGATCGCGCGCGGCA 1166
QY 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
DB 1167 CGAAGCGAGTTCGGCGAGCAGACCATCGGCTCCCGCGGCGCGCGCGCGCGCTCAC 1226
QY 180 rValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHi 200
DB 1227 CGTCGCGCGC-----GTGAACGACGACGACAA 1253
QY 200 sileAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220

```

```

DB 1254 GCTGGCTCTCTCCAGCGCGCGCGCGCTGGAGCGCGCATCAAGCCGCGCTCAC 1313
QY 220 xAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
DB 1314 CGACCCGCGGTGGACATCACCGCGC-----GCGGCGCGCGCGCGCTCATCGA 1364
QY 240 pAlaAsnTyrAsnSerLys-----TyrAlaTyrMetGlyGlyThrSerMetAlaTh 257
DB 1365 CCAGGAGGTGCGGCGGAGCGGCTACCTCACCATCTCCGTACGTGATCGCGCGAC 1424
QY 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI 277
DB 1425 CCGCATGTGCGGCGCGCGCGCGATCTCAGCAGCAGCAC----- 1467
QY 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
DB 1468 ----CCAACTGTGTCTTCGCGAGCTCAAGGCGCGCTGACCGGCTCCGCGAAG----- 1518
QY 295 lGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLe 315
DB 1519 -GCGGCAAGTACACCGCGTTCAGCAGGCGTCCGCGCGCTATCGCGCTGACAGCGCAT 1577
QY 315 uAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSe 335
DB 1578 CAAGCAGTCCGTGATCGCAACCGCAACTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGC 1636
QY 335 rPheGlnAlaGlnAlaGlyLysPro 343
DB 1637 ACACCGCAGCAAGCGCGTCAACCCA 1661

RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.37e-20 Length: 9025608
Score: 329.00 Matches: 114
Percent Similarity: 45.27% Conservative: 44
Best Local Similarity: 32.66% Mismatches: 126
Query Match: 14.55% Indels: 67
DB: 14 Gaps: 12

US-09-985-689A-5 (1-433) x US-10-156-761-1 (1-9025608)
QY 18 GlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37

```



```
Db 4132505 GGGTACAGCGCAGGGCGTCAAGATCGCGTCTCGACACCGGTCTGCAC----- 4132455
Qy 38 AsnAspSerMetHisGluAlaPheArgGlyLeuThrAlaLeuValLeuVal 57
Db 4132454 -----CGACACCGCGGACCTCAAGACCGAGTGGCCGAGTCCAAAGACTTCTCC 4132404
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 4132403 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4132344
Qy 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 4132343 GGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4132284
Qy 93 PheGlnSerIleMetAspSerGly-----GlyLeu 103
Db 4132283 AACGGCAAGTCTTCGACGACCGCGTCCGGCGACGATCCGGCGCGCGCGCGCG 4132224
Qy 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 4132223 GAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4132164
Qy 123 GileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgG 143
Db 4132163 CCGGAGATCGACCGCGCTGGAA-GCGGAGGTCAACAAGCTC----- 4132125
Qy 143 nValAspGluTrpValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 4132124 -----TCCGAGGAGAGGGCGATCTCTTCGCGATCGCGCGCGCGCAA 4132084
Qy 161 nGlyGlyPro--AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLeu 180
Db 4132083 CGAAGGCGAGTTCGCGGACGACGATCGGTCTCCCGCGGCGCGCGCGCGCGCG 4132024
Qy 180 rValGlyAlaThrGluAsnTrpArgProSerPheGlySerLeuAlaAspAsnProAsn 200
Db 4132023 CGTGGCGGCC-----GTGAACGACGACGCGCAA 4131997
Qy 200 sIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspVal 220
Db 4131996 GCTGGCGTCTCTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4131937
Qy 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTr 240
Db 4131936 CGCACCGCGGTGGACATCAACGCGCC-----CGCGCGCGCGCGCGCGCGCG 4131896
Qy 240 pAlaAsnTrpAsnSerLys-----TyrAlaTrpMetGlyGlyThrSerMetAla 257
Db 4131895 CCAGGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4131826
Qy 257 rProfileValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGly 277
Db 4131825 CCGGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4131783
Qy 277 eThrProLysProSerLeu-----IleLysAlaAlaLeuIleAlaGlyAlaThr 295
Db 4131782 -----CCCAACTGTCGTCGCGAGCTCAAGGGCGCGTGCACCGCGTCCGCG 4131732
Qy 295 lGlyLeuGlyTrpProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSer 315
Db 4131731 -GGCGCGCAAGTACACCGCGTCCAGCGGCTCGCGCGCGTATCGCGCGTCAAGAGG 4131673
Qy 315 uAsnValAlaTrpValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTrp 335
Db 4131672 CAAGCAGTCCGCGTATCGCGCAACCGCACTCGGTGAGTTCGGCAT-CCAGCAGT 4131614
Qy 335 rPheGlnAlaGlnAlaGlyLysPro 343
Db 4131613 ACACCGACGACAGCGCGGTCAACCA 4131589
```

RESULT 9

US-10-156-761-5701

```
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
; US-10-156-761-5701
```

```
Alignment Scores:
Pred. No.: 2,71e-25 Length: 3624
Score: 327.00 Matches: 119
Percent Similarity: 42.75% Conservative: 49
Best Local Similarity: 30.28% Mismatches: 161
Query Match: 14.4% Indels: 64
DB: 16 Gaps: 16
```

US-09-985-689A-5 (1-433) x US-10-156-761-5701 (1-3624)

```
Qy 8 ValLysAlaAspValAlaGlnAsnAsn-----Tyr 17
Db 556 GTGAGGCGGACATCGCGAGGACACCGGAGATCGGTACGCGCGCGCGCGCGCG 615
Qy 18 GlyLeuTrpGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGGCTCACGGGCGAGCGGGGTCACTCGTCCGCGTGTGACACCGCGCGTGCACAC 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTrpAlaLeuGly 57
Db 670 -----ACTCACCGACCTCGCGCGCGGGTGTCCCGGAGCAAGGCTTCATC 717
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 GACGGGAGGAGGTGCGCGACCGCAACCGGACCGGACCGGACCGTCCCTCGATCGCG 777
Qy 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 778 GGCAGCGCGCGCGTTCACGCGGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 837
Qy 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 838 GTGCGCAAGTGTCTACGACCGACCGAGGCGCG-----GGAAGCGAGTCCAGATC 885
Qy 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaAspGileHisThrAsnSerTrp 129
Db 886 ATCGCGCGCATGAATGGCGCGCGGAGCGGTGCGCGGAGATCGTCTCGATGAGCCTC 945
Qy 130 GlyAla---ProValAsnGlyAlaTrpThrAlaAsnSerArgGlnValAspGluTrpVal 148
Db 946 GGATCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1005
Qy 149 ArgAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr 168
Db 1006 GAGGAGACCGCGCGCGCTCTTCGTGTCGCGCGGGAACACCGGTGCGCGCGCTCC 1062
```


QY 288 -----LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProSerGlyAspGingly 305
Db 6919626 GAGCAACTCAGCGCTCCGATATATCATAGTTCGGGGCGGTCGGTCAAGTTCGGGAGCGC 6919685
QY 306 TrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr-----ValAsn 321
Db 6919686 GTGGCGCGCGCTCACCGGAGCGGCGGCGGCGGCTTCACCGCTGCCGCCAT 6919745
QY 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
Db 6919746 GAGCCGATCGACCGTCACG-----AAGCGGTACCTTACTTCCAACTCCCTCCGAC 6919796
QY 342 LysProLeuLysLysSerLeuValTrpThrAspAlaProGlySerThrThrAlaSerTyr 361
Db 6919797 ACGACGTCGAGTTCAGCTCGCGTCCGGCGGCGGCGGCGGCTGCGCC----- 6919847
QY 362 ThrLeuValAsnAspLeuValLysAlaPro 374
Db 6919848 ACCCTCGCC-----GACACCGCACTCACCGTGCCC 6919877

RESULT 11

US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Miko
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyoze
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:

Pred. No.: 6,41e-23 Length: 4765
Score: 307.00 Matches: 129
Percent Similarity: 36.48% Conservative: 37
Best Local Similarity: 28.35% Mismatches: 153
Query Match: 13.58% Indels: 136
DB: 13 Gaps: 15

US-09-985-689A-5 (1-433) x US-10-090-624-5 (1-4765)

QY 21 GlyGinglyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyValGAsnAspSer 40
Db 1024 GCAATGTTATGACATTGCATATGTGTGATACCTTGACTGACCTTGCCTCAGCTTCCCGACGAA 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
Db 1084 GTTCCACTTGGCCAGTACACGTTACTTATGATGTTGCTGTTTTTAGCTACTACTACGGT 1143
QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn----- 66
Db 1144 CCTCTCAAC-----TAGTGTCTCCAGAAATA-----GATCCTTACGGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1186 TATGAGTATTGGTGGGATGGTTCACGGTACCGAACTACGAGTACCTGGAACCTGTTGCT 1245

QY 78 GlyAsnAlaLeuAsn----- 82
Db 1246 GGTACGACGACCAATGATGCTTGGATTGGCTCAGTATGTAATGGAA 1305
QY 83 -----LysGly 84
Db 1306 GTGTTCTCAGACTCTATGTTGGATTATACGAACGTTTACCACAGACACCGTCACGGT 1365
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
Db 1366 GTTCTCAGGTGCGCAAAATGCAATAGAGTCTTAGAGTGTATGA-----CGG 1419
QY 105 GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIle 124
Db 1420 GGTAGCATGTGGGATATTATAGAAGGTATGACATACGCAACCCATGTTGTCAGACGTT 1479
QY 125 HisThrAsnSerTrpGly-----AlaProValAsnGlyAlaTyrThrAlaAsnSerArg 142
Db 1480 ATAAGCATGAGTCTCGGTGGAAATGCTCCATCTAGTGTACTGATCCAGAAAGCGTT 1539
QY 143 GlnValAspGluTyrValArgAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
Db 1540 GCTGTGGATGAGTTCACGAAAGTACGGTGTGTTGTTATTCGTAATAGTTCGAGGAATGAA 1599
QY 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 1600 GGTCTCTGCATTAACATCGTTTGAAGTCTCTGTTGTCACAAAGGCAATACCTGTGGA 1659
QY 183 AlaThrGlu----- 185
Db 1660 GCTGCTGCAGTGCCTCCATTACGTTGGAGTTTATGTTTCCCAAGCACTGGGATATCTGAT 1719
QY 186 -----AsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHisIle 201
Db 1720 TACTATGATCTATTACTTCCCGCTTACACAAACGTT-----AGAATA 1764
QY 202 AlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAla 221
Db 1765 GCATTCTTCTCAAGCAGAGGCGCGAGATAGATGGTGAATAAAACCAACCAATGATGGCT 1824
QY 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
Db 1825 CCAGGTTACGGAATTTACTCATCTCCGCGGATGGATGGCGGAGCTGACTTC----- 1878
QY 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 261
Db 1879 -----ATGTCGGAATTCGATGGCTACTCCACATGTTCAGC 1914
QY 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
Db 1915 GGTGTGCTTGCATCTCTCATAGCGGG-----GCAAGGCGCGGGAATATATACATCCA 1971
QY 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThr-----AspVal 295
Db 1972 GATATAATTAAAGGTTCTTGAGAGCGGTGCACCTGGCTTGAGGAGATCCATATACT 2031
QY 296 GlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLeuAspLysSerLeu 315
Db 2032 GGGCAGAAGTACACTGAGCTTGACCAAGTCTGTTGTTTAACTGTTTAACTGTTTAACTGTT 2091
QY 316 AsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSer 335
Db 2092 GAATTC----- 2097
QY 336 PheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGly 355
Db 2098 CTTAAGGCTATAAAGCGCACCACTCTCCCAATTGTTGATCACTGGCGCAGACCAAGTCTTAC 2157
QY 356 SerThrThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsn 375
Db 2158 AGCGACTTTGGCGAGTACTTGGGTGG-----GACGTTATAGAGGTTCTTACGCAAGGAC 2214


```
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 2,66e-22 Length: 1560
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservatives: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-113 (1-1560)
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTACGGCGTACGTCATCGACACCGCGCTC-----CGC 492
Qy 41 SerMetHisGluAlaPheArgGlyLeuThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCGATTCGGCGCGCGCGCTCTCTACCGCTACGACGCGTCAACACGAC 552
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCCGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
Qy 80 AlaLeuAsnTyrGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 613 GCCTAC-----GGCGTCGCGCAAGAGCAAGATCGTAGCGTCCGCGTCTGTAACAC 666
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 667 TCCGGCGAGGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 727 GT-CAAGCGCGCGTCCGCAACAT-----GTCCCTCGCGCGCGCGCGCGCGCG 770
Qy 137 rThrAlaAsnSerArgGlnValAlaGlyTyrValArgAsn-----AsnAspMetTh 154
Db 771 CACGCGC-----CTCGACACGCGCGTACGACACCGCATCGCGTCCGCGCGCGCGCG 818
Qy 154 rValLeuPheAlaAlaGlyAsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 819 CTTCGCGTGGCGCGCGCAACGAGTCGACCAACGCTCCACGAGGTCA---CCCGCACG 875
Qy 174 rAlaLysAsnAlaLeuThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLe 194
Db 876 CGTCACCGAGGCGCATCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930
Qy 194 uAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 931 -----TACTCCAACTACGGCTCCGCTCCTC-----954
Qy 214 GileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 955 -----GACCTCTTCGCGCGCGCGTCTGCTTCCATCCTCGCGC-----990
Qy 234 aproAspSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 991 -----TGAACCTACGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1034

; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 2,66e-22 Length: 1560
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservatives: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-113 (1-1560)
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTACGGCGTACGTCATCGACACCGCGCTC-----CGC 492
Qy 41 SerMetHisGluAlaPheArgGlyLeuThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCGATTCGGCGCGCGCGCTCTCTACCGCTACGACGCGTCAACACGAC 552
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCCGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
Qy 80 AlaLeuAsnTyrGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 613 GCCTAC-----GGCGTCGCGCAAGAGCAAGATCGTAGCGTCCGCGTCTGTAACAC 666
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 667 TCCGGCGAGGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 727 GT-CAAGCGCGCGTCCGCAACAT-----GTCCCTCGCGCGCGCGCGCGCGCG 770
Qy 137 rThrAlaAsnSerArgGlnValAlaGlyTyrValArgAsn-----AsnAspMetTh 154
Db 771 CACGCGC-----CTCGACACGCGCGTACGACACCGCATCGCGTCCGCGCGCGCGCG 818
Qy 154 rValLeuPheAlaAlaGlyAsnGlyGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 819 CTTCGCGTGGCGCGCGCAACGAGTCGACCAACGCTCCACGAGGTCA---CCCGCACG 875
Qy 174 rAlaLysAsnAlaLeuThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLe 194
Db 876 CGTCACCGAGGCGCATCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 930
Qy 194 uAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 931 -----TACTCCAACTACGGCTCCGCTCCTC-----954
Qy 214 GileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 955 -----GACCTCTTCGCGCGCGCGTCTGCTTCCATCCTCGCGC-----990
Qy 234 aproAspSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 991 -----TGAACCTACGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1034

254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
1035 GATGCGACCCCGCAGCTGCGCGCGCGCGCGCTC-----CACCTCGCGCGCAA 1085
274 nArgGlyIleThrProLys-----ProSerLeuIleLeuAlaLeuIleAlaG 291
1086 CCCCTCGGCCACCCCGTCCGCGCGCGCGCTGACGTCCGCGCGCGCACCCACCGCGCT 1145
291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
1146 CGTCACCAACCCCGCGCGCGCTCGCCCAAC-----1176
311 uAspLysSerLeuAsnValAlaValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
1177 -----CGGCTCCTGTAGTCCGCGCGCGCGCACCCACCTCCGCGCGCGCG 1223
331 s-----AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLy 345
1224 CTTGAGAACACCGCTGACTACAGTACGACGACCACTCCACGCTCGAGTCCCGGTGAC 1283
345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
1284 GGTCTCCGCGCTCTCCGCGCAACGCGCTCGGCCCTCGCGCTAGAGTCCACATCGTCCA 1343
361 rThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGl 381
1344 CACGTACATCGCGACCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 1403
381 yAsnAspPheSerTyrProTyrAspAsnAsnTyrAspGlyArgAsnValGluAsnVa 401
1404 CAAG-----TCGTACGCGCACCGCGCGCGTCCGCAACATCAACACCATCGTACTCGGT 1457
401 lPheIleAsnAlaProGln-----SerGlyThrTyrThrIleGluVal 415
1458 G-----AACGCTCTCTCGAGCGCGCGCACCGCGCGTGGAAACTGGCGGTG 1503

RESULT 14
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNBERG, AGNES
; APPLICANT: TREPZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: PCT/EP01/09815
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

Alignment Scores:
Pred. No.: 5.25e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservatives: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-1 (1-59816)
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
```

Db 56648 GGGCAGGAGTACGGCGTACGTCTCGACACGGCGTC-----CGC 56689
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 56690 ATCACCCACAGCGACTTCGGCGCGCGGCTCTCTACGGGTACGACGCCATCGACACGAC 56749
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACGCCAGCAGCGCCGCGCCACGCGCAGCGACGTGGCGGCGGCGGTGCGCGCAAC 56809
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 56810 GCCTAC-----GGCGTCGCCAAGAGGCCAAGATCGTAGGGTCCGGGTCTGAACAAC 56863
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 56864 TCCGGCCAGCGCACCCCGCCAGTCTCGCGCGCATCGACTGGGTCCCGCGGACGCC 56923
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 56924 GT-CAAGCGCGCGTCCGCAACAT-----GTCCCTCGGGCGGCGCGCA 56967
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 56968 CACGGCC-----CTCGACACGGCGGTACGACCAACGCGATGGCTCCGGGTGCAC 57015
Qy 154 rValLeuPheAlaAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 57016 CTTCCGGCGTGGCGCGCGCAACAGTCCGACCAAGCTCCAGAGTCA---CCGCGACG 57072
Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLe 194
Db 57073 CGTCACCGAGGCATCAGCGTCGGCGCGACGACCGACGCGCGCGCGC----- 57127
Qy 194 uAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 57128 -----TACTCCAACTACGGCTCCGTCCTC----- 57151
Qy 214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 57152 -----GACCTCTTCGCCCCCGTTCGTCTCATCACCCTCGGC----- 57187
Qy 234 aproAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSe 254
Db 57188 -----TGGAACTCAAGCGACTCGGCGGACCAACACCATCTCCGGTACGTC 57231
Qy 254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db 57232 GATGGCAGCCCGCACGTGGCGGCGCGCGCTC-----CACCTCGCGCCAA 57282
Qy 274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaG 291
Db 57283 CCCCCTCGGCCACCCCGTCCAGGTCGCCAGCGGCTGACGTCCGCGCCACACCGCGGT 57342
Qy 291 yAlaThrAspValGlyLeuGlyTyrProSerGlyAspGlnGlyTrpGlyArgValThrLe 311
Db 57343 CGTCACCAACCCCGGACGGGCTCGCCCAAC----- 57373
Qy 311 uAspLysSerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
Db 57374 -----CGGCTCTCTGTACGTCCGCGCGCGCGCACACCCCTCCGGCGCGCGG 57420
Qy 331 s-----AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLy 345
Db 57421 CTTCCGAGAACACCGGTGACTACATCAGCAGCACTCCACGGTCCAGTCCCGGTGAC 57480
Qy 345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 57481 GGTCTCGGCGTCTCCGCGCAACGGCGCTCGCGCCCTCGCGGTAGAGTCCACATCGTCCA 57540
Qy 361 rThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrValG 381
Db 57541 CACGTATACGGCACCTCCAGGTCACGCTGATCGCCCCCGACGCGCGCGGTACAGCT 57600

Qy 381 yAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 57601 CAAG-----TCGTACGGCACCGCGCGCATTCGGACAACATCAACACACGCTACTCGGT 57654
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrThrIleGluVal 415
Db 57655 G-----AACGCCTCTCGGAGCGGCGCACGCGACGTGGAAACTCGCGGTG 57700
RESULT 15
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIORITY FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 5,25e-20 Length: 59816
Score: 295.00 Matches: 125
Percent Similarity: 42.11% Conservative: 51
Best Local Similarity: 29.90% Mismatches: 153
Query Match: 13.05% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-5 (1-433) x US-10-084-846A-2 (1-59816)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTGACGGCGGTACGTATCGACACCGCGCTC-----CGC 3128
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 3127 ATCACCCACAGCGACTTCGGCGCGCGCTCTCTACGGTACGACGCGCATCGACACGAC 3068
Qy 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACGCCAGGACGCGCCACGCGCACGCGCACGCGTGGCGGCGCGGTGCGCGCAAC 3008
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 3007 GCCTAC-----GGCGTCGCCAAGAGGCCAAGATCGTAGGGTCCGGTCTGAACAAC 2954
Qy 100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 2953 TCCGCGCAGGCGCACCCACCGCGCGCATCGTCCGCGCATCGACTGGGTGCGCGGACGCC 2894
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 2893 GT-CAAGCGCGCGTCCGCAACAT-----GTCCCTCGGGCGGCGCGCA 2850
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 2849 CACGGCC-----CTCGACACGGCGGTACGCAACGCGCATGGCTCCGGGTGCAC 2802
Qy 154 rValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyTh 174

Mon Apr 5 16:03:39 2004

[illegible]

Search completed: April 5, 2004, 00:57:57
Job time : 6868.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-5
Perfect score: 2261
Sequence: 1 NDVARGIVKADVAQNNGLY.....EVQAYNVSPQSFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q/cgn2_1/USPTO.spool/US09985689/runat_31032004_161807_4168/app_query.fasta_1.3498
-DS=EST -OFMT=fasta -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN_1_1_7084 @runat_31032004_161807_4168 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	235.5	10.4	1605	13	BQ622771	BQ622771 CC Contig
2	228	10.1	4198	11	AK029048	AK029048 Mus muscu
3	215.5	9.5	594	12	BQ393752	BQ393752 BQ393752
4	214.5	9.5	640	12	BQ395336	BQ395336 BQ395336
5	204	9.0	2141	13	BQ142519	BQ142519 Contig M
6	199.5	8.8	532	29	TA319G10P	AL492464 T. brucei
7	199.5	8.8	718	12	BI750157	BI750157 Fg02.10G0
8	195.5	8.6	771	14	CA320325	CA320325 UI-M-FW0-
9	193.5	8.6	601	12	BQ387574	BQ387574 BQ387574
10	190	8.4	545	13	BQ575479	BQ575479 T9ESTZyb8
11	188	8.3	574	29	TA315H10P	AL490202 T. brucei
12	188	8.3	614	9	AJ273402	AJ273402 AJ273402
13	183.5	8.1	2121	28	BZ424955	BZ424955 100023066
14	182.5	8.1	665	13	BQ770462	BQ770462 UI-M-FI0-
15	182.5	8.1	716	28	BZ893395	BZ893395 HL2.0177
16	180	8.0	530	29	CNSC10FO	AL153820 Anopheles
17	180	8.0	650	9	AJ274038	AJ274038 AJ274038
18	179.5	7.9	675	14	CF727824	CF727824 UI-M-HB0-
19	178	7.9	508	28	AQ652212	AQ652212 Sheared D
20	178	7.9	633	12	BQ369190	BQ369190 BQ369190
21	177.5	7.9	576	14	CD295943	CD295943 StrPu691.
22	176.5	7.9	1002	29	CNS0606B	AL393417 T3 end of
23	176.5	7.8	594	14	CF846026	CF846026 psHB036XA
24	176.5	7.8	1029	29	CNS071DW	AL424794 T7 end of
25	175	7.7	580	9	AJ273745	AJ273745 AJ273745
26	175	7.7	583	9	AJ273947	AJ273947 AJ273947
27	175	7.7	593	9	AJ273918	AJ273918 AJ273918
28	175	7.7	601	9	AJ273050	AJ273050 AJ273050
29	175	7.7	601	9	AJ273921	AJ273921 AJ273921
30	173.5	7.7	641	12	BQ393925	BQ393925 BQ393925
31	172	7.6	610	9	AJ274218	AJ274218 AJ274218
32	170	7.5	564	28	AQ651427	AQ651427 Sheared D
33	169.5	7.5	937	12	BQ246418	BQ246418 602360428
34	169	7.5	609	9	AJ272712	AJ272712 AJ272712
35	169	7.5	794	14	CF737198	CF737198 UI-M-HD0-
36	168	7.4	781	13	BQ612128	BQ612128 UI-M-EW0-
37	167.5	7.4	604	9	AJ274126	AJ274126 AJ274126
38	167.5	7.4	615	9	AJ274059	AJ274059 AJ274059
39	167	7.4	616	13	CF138007	CF138007 UI-HF-BN0
40	166.5	7.4	895	13	BQ216158	BQ216158 AGENCOURT
41	166	7.3	619	9	AJ273097	AJ273097 AJ273097
42	166	7.3	681	14	CB690041	CB690041 CBST-54-B
43	166	7.3	887	13	BQ879057	BQ879057 AGENCOURT
44	165.5	7.3	593	9	AJ273903	AJ273903 AJ273903
45	165.5	7.3	596	9	AJ273379	AJ273379 AJ273379

ALIGNMENTS

RESULT 1
BQ622771

LOCUS

DEFINITION

CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus

CDNA, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1605)

BQ622771 1605 bp mRNA linear EST 01-JUL-2002
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.

BQ622771.1 GI:21649940

EST.

Conidiobolus coronatus (Delacroixia coronata)

Conidiobolus coronatus

Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;

Ancylidiaceae; Conidiobolus.

1 (bases 1 to 1605)

AUTHORS

Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
EST analysis of genes expressed by the zygomycete pathogen
Conidiobolus coronatus during optimized secretion of proteins
Unpublished (2002)

JOURNAL

Contact: Freimoser F. M.

Department of Entomology

University of Maryland

4112 Plant Sciences Building, College Park, MD 20742, USA

Tel: 301 405 16 13

Fax: 301 314 92 90

Email: ff34@umail.umd.edu.

Location/Qualifiers

1. .1605

/organism="Conidiobolus coronatus"

/mol_type="rRNA"

/strain="ARSEF 512"

/db_xref="taxon:34488"

/clone_lib="Conidiobolus cornatus ARSEF 512"

/note="Vector: UniZap; Conidiobolus coronatus was grown in
minimal medium supplemented with Manduca sexta cuticle and
peptone for 18 hours. A cDNA library was constructed in
the unidirectional Lambda vector UniZap."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 1.07e-12 Length: 1605
Score: 235.50 Matches: 88
Percent Similarity: 43.40% Conservative: 37
Best Local Similarity: 30.56% Mismatches: 88
Query Match: 10.42% Indels: 75
DB: 13 Gaps: 14

US-09-985-689A-5 (1-433) x BQ622771 (1-1605)

Qy 6 GlyIleValIysAlaAspValAlaGlnAsnAsn-----TyrGlyLeu----- 19
Db 718 GGGTCTAGGCTATTGCGGTCACAGCAATGCTTCCTGGGCTTCTGCTGGTGT 777
Qy 20 -----TyrGlyGlnGly 23
Db 778 CAACGTCTAAGCTGGATCTGCTTACTTACTTACCAACCATCTGCTGCTCAAGT 837
Qy 24 GlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHis 43
Db 938 GTCACTGTTTTCGTTTGTAGTACTGGTGTCAATGTCAGCCACCAATGAC----- 885
Qy 44 GluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAla---- 62
Db 886 -----TTCGTGTCTGCGCCACT-----TGGGTACTAACACTGCTGCTGT 924
Qy 63 -----AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 925 GGTAGCAACACTGATGCTCAGGTCACGGTACTCAGTCTGCTGCTATTTGCTGTACC 984
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet----- 97
Db 985 ACCTAT-----GGTGTGTCACAGAGCTAACATTTCTCCGTAAAGGCTTAGGTGAT 1038
Qy 98 AspSerSerGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 1039 GATCGCTCCGGATCACTACTCGGAATTATCTCCGGTATTCAGTGGTGTAAAGCAC--- 1095
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTy 137
Db 1096 ---TCTGCTCCAGAAAGTTATCTCTATGATTAGAGGTGTAAGACGATGCTCTT 1152
Qy 138 ThrAlaAsnSerArgGlnValAspGluTyValArgAsnAsnAspMetThrValLeuPhe 157
Db 1153 -----AACACTGCTGTTTAAACACGCTGTCAGCAGGAGGTGTCAGTGTGTC--- 1200
Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsn 177
Db 1201 GCTGTGTAAAGTAAACAGAGATGCTGTGGTACTCT---CCCGCTTCTGCTCTTCC 1257

Qy 178 AlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySerLeuAlaAspAsn 197
Db 1258 GCCATTACCGTTGGTCCACT-----GATGTC 1284
Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysPro 217
Db 1285 AATGATAAAAGGCTTCATTCTCTAACTCGTAGTGT-----GTC 1326
Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
Db 1327 GATATCTTAGCTCTCTGCTGTCACATCTCTCCACC----- 1362
Qy 238 SerPheTrpAlaAsnTyArgSerLysTyAlaTyMetGlyGlyThrSerMetAlaThr 257
Db 1363 -----TGAAGGATCTAACTGACCAACCATCTCTGTTACCTTATGGTGTGC 1416
Qy 258 ProIleValAlaGlyAsnValAla 265
Db 1417 CCTCACATTGCTGTTAGTGTCT 1440
RESULT 2
AK029048
LOCUS
DEFINITION
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
library, clone:473248M11 product:membrane-bound transcription
factor protease, site 1, full insert sequence.
ACCESSION
AK029048
VERSION
AK029048.1 GI:26325017
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349836
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
6 (bases 1 to 4198)


```

QY 314 SerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThr 333
D 1827 GCTATCAGATCTC-----AGCAGC 1847
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeu----- 348
D 1848 TATAAACCGCAGCAAGCCTG---AGTCCTAGCTACATCGACCTGACTGAGTGCTCCCTAC 1904
QY 349 ValTyrThrAspAlaProGlySerThrThrAlaSerTyr-----ThrLeu 363
D 1905 ATGTGG-----CCCTACTCTCCAGCCTATCTACTATGGAGGAATGCACAAATC 1955
QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAsp 383
D 1956 GTTAAT-----GTCCATCCTCATATGGCATGGCGCTCACAGGAAGAAT 2000
QY 384 PheSerTyrProTyrAspAsnAsnTyr-----AspGlyArgAsnAsnVal 398
D 2001 GTGGATAAGCCT-----GAGTGGCGACCCCTATTTACCACAGNAATGGAGACAAAT 2051
QY 399 GluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyrThrIleGlu 414
D 2052 GAATGGCCCTTCTCCTACTCTCCTCAGTGTGTGGCCCTGTGTGAGTTACCTTGCCATCTCC 2111
QY 415 Val 415
D 2112 ATT 2114

RESULT 3
LOCUS BJ393752
DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BJ393752
VERSION BJ393752.1 GI:19304838
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 594)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..594
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores:
Pred. No.: 2,13e-11 Length: 594
Score: 215.50 Matches: 55
Percent Similarity: 53.18% Conservative: 37
Best Local Similarity: 31.79% Mismatches: 60
Query Match: 9.53% Indels: 21
DB: 12 Gaps: 8

US-09-985-689A-5 (1-433) x BJ393752 (1-594)

```

```

QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProHsnGlyHis 68
D 21 AAAGTTGTAATCTATATTACCATCAACAAAGCGAGATAGTGAAGTGGATGGTGCAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGly-----AsnAlaLeuAsn----- 82
D 81 GGTACACATATTTGTGGTTCTGCAGAGAGTACTCCAGAGGATTTCTTCAGTTATATTTCA 140
QY 83 -----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
D 141 TCATTTAGTGGTCTTGCACACTGATGCAAGATTCATTTCTTTGATTGGCAAGTGGTCA 200
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrAsnAla 120
D 201 TCAAGTTTGACACCT---CCATCGAATTTGAACAATTTATCAACATATATATGAGCA 257
QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAla-----ProValAsnGlyAla 136
D 258 GGTCGAAGAGTGCATTCTGATTCTTGGGTTCTGTATCAGTAGAGGGGTATACAGGTAGT 317
QY 137 TyrThrAlaAsnSerArgGlnValAspGluTyrVal---ArgAsnAsnAspMetThrVal 155
D 318 TATTCATCAGACACTGTTCAATTTGATGATTTCTTTCTCACTCATCCAGATTTCACTTT 377
QY 156 LeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSer-----AlaProGly 173
D 378 CTTAGAGCTGCTGTTAAC-----AACGAGCAATACCTATCACTCACTCAATCC 428
QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySer 193
D 429 ACTCAAGAGATGTTATTACCGTTGGTGCTCATCAACAATTCATGAAAATTAAT----- 482
QY 194 LeuAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
D 483 TTAACGTGATGCTCCAAATATATATAATATCAATCATCT 521

RESULT 4
LOCUS BJ395336
DEFINITION BJ395336 Dictyostelium discoideum cDNA library, EST 08-MAR-2002
discoideum cDNA clone dds38b16 5', mRNA sequence.
ACCESSION BJ395336
VERSION BJ395336.1 GI:19306422
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 640)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..640
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores:
Pred. No.: 3,06e-11 Length: 640
Score: 214.50 Matches: 67

```

Percent Similarity:	50.47%	Conservative:	41
Best Local Similarity:	31.31%	Mismatches:	65
Query Match:	9.49%	Indels:	41
DB:	12	Gaps:	12
 US-09-985-689A-5 (1-433) x BJ935336 (1-640)			
QY	19	LeuTyrGlyGlnGlyClnValValAlaValAlaLalaspThrGlyLeuAspThrGlyArg---	37
Db	22	TTAAGAGTAAAGGTTCAGATATTGGATATTCTGTACTTGTTTAGATCGTAGCCATTGT	81
QY	38	-----AsnAspSer-----SerMetHisGluAlaPheArgGly	48
Db	82	TTCCTTTTCAGATCAAGTATCCATACCATTTTAATCAAGTGAATGAATAATCATAGAAA	141
QY	49	LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro-----	Asn 66
Db	142	GTGTAACTTATATT-----ACTTACCATGACAATGAAGAATTATGTAAAT	186
QY	67	GlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-----	Ala 80
Db	187	GGTCATGGTACACATGTTTGTGGCTCTCCAGCAGGTACTCCAGAGGATTCCTCATGGCT	246
QY	81	LeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp	98
Db	247	ATTTCATCATTTAGTGGTCTTGCAACTCACGCCAAGATTGCATTTTATGATCTT----	300
QY	99	SerSerGlyGlyLeuGlyClyLeu---ProSerAsnLeuAsnThrLeuPheSerGlnAla	117
Db	301	TCATCTGGAAGTCTTGAACCAACACCCCAGGAAGTTACAGTCAAATGTACAAACCATTA	360
QY	118	TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----	ProVal 133
Db	361	TATGATGCAGGTGCAAGAGTACATGGTGATCTTGGGGTCTGTATCTTTGCAAGTTAT	420
QY	134	AsnGlyAlaTyrThrAlaAsnSerArgGlnValAlaspGluTyrVal--ArgAsnAsnAsp	152
Db	421	TATGGTGGTATTCCCATGCTCGTGGTATTCATGTCATCTCTATAGTACCAGAA	480
QY	153	MetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlap	172
Db	481	TTCTCTATACTAGAGCTGCTGGTAT--AACGAGCTATTGGCATCTTTATTAGCTCAA	537
QY	173	GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGly	192
Db	538	GCAACAGCTAAAAATGCAATTACAGTTGGTGTGAGCAACACAGCTCATGTAAATTTATG	597
QY	193	Ser-----LeuAlaAspAsnProasn	199
Db	598	TCAGATGCATTTGAATATTATGATTTCTCAGATAATGCTAAT	639
 RESULT 5 BQ142519/c LOCUS DEFINITION Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium anisopliae var. acridum cDNA, mRNA sequence. ACCESSION VERSION KEYWORDS SOURCE EST. ORGANISM Metarhizium anisopliae var. acridum Metarhizium anisopliae var. acridum Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium. REFERENCE AUTHORS TITLE JOURNAL COMMENT			

```

Qy 255 MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgLuhHisPheIleLysAsn 274
Db 471 ATGGCTACTCCCATATGCTGT-----CTTGCTGCCCTACTTCAGTCTCTC 424
Qy 275 ArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAsp 294
Db 423 AGCGGCAAGACTAGCCCTGCGTCTTTGCCAGAGATCCAGGACACTTCTACCAAGAAC 364
Qy 295 ValGlyLeuGlyTyrProSerGly 302
Db 363 GTGATCCGCAATGTGCCCGCTGGC 340

RESULT 6
TA319G10P 532 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492464
VERSION AL492464.1 GI:11867408
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 532)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nh@sanger.ac.uk
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES
source
1..532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319G10"

ORIGIN
Alignment Scores:
Pred. No.: 7,53e-10 Length: 532
Score: 199.50 Matches: 59
Percent Similarity: 46.74% Conservative: 27
Best Local Similarity: 32.07% Mismatches: 71
Query Match: 8.82% Indels: 27
DB: 29 Gaps: 6

US-09-985-689A-5 (1-433) x TA319G10P (1-532)

Qy 92 ValPheGlnSerIleMetAspSerSerClyGlyLeuGly----- 104
Db 2 GTAATGCCAAGTATAATGTTCTCCAGGGGCGGAGATTCTTCAGGGTGGCGTGC 61
Qy 105 -----GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 62 CATCCAGTCAGAGCTTGCTCTCCGCCACGACGTTACTCAAAATTATTCGTCGCGGTAT 121

```

```

Qy 119 AsnAlaGlyAlaArgIleHisThrAsnSerTirpGlyAlaProValAsnGlyAlaTrpThr 138
Db 122 GGCCCTGGAGCCCGTGTTCTTCAACTCGTGGGTTTTTGTCTCCTCCGAGTATTCT 181
Qy 139 AlaAsnSerArgGlnValAspGluTyrValArgAsn---AsnAspMetThrValLeuPhe 157
Db 182 GCTGTGGAAGAGATATGATGAGTTTTCGAGTAGTATTATGACGATGCGCTACTTATCTTC 241
Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLeuAsn 177
Db 242 TCCACTGGCACAGTATCCAGATGGC-----CTAATGACTCGTGTCTGTGTAAGAAC 295
Qy 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsn 197
Db 296 GTGATGTGCGTGGGCTCACACAAAACGTG-----TTTGACGCTTCCGAAAGAC--- 343
Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysPro 217
Db 344 -----ATTGTTTCTTCGTTTCTTCGATGGTCCACATACACGCTAGATGAACCC 397
Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
Db 398 GATCTGTGCGTCCCGGGGAGAGGTGTGTCTCTTCTTCTGCGCAAGCATCAGCT 457
Qy 238 SerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThr 257
Db 458 AAA-----CAATGTAAAGTGGTGCCCAAGCGGNGTTTCATCGATGGCAACT 502
Qy 258 ProIleValAla 261
Db 503 CGCGCGCTCGCG 514

RESULT 7
BI750157 718 bp mRNA linear EST 25-SEP-2001
LOCUS Fg02_10g08 R Fg02_AAFCECORC Fusarium graminearum mycelium
DEFINITION Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
ACCESSION BI750157
VERSION BI750157.1 GI:15771959
KEYWORDS EST.
SOURCE Gibberella zeae
ORGANISM Gibberella zeae
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocerales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 718)
AUTHORS Harris,L.J., Glasco,T., Rocheleau,H., Allard,S., Chapados,J.,
Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S.,
Singh,J.A., Spratt,D. and Finkler,N.A.
Expressed Sequence Tags from Fusarium graminearum mycelium
Unpublished (2001)
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, KIA 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harrisli@em.agr.ca.
Location/Qualifiers
1..718
/organism="Gibberella zeae"
/mol_type="mRNA"
/strain="DAOM 180378"
/db_xref="taxon:5518"
/clone="Fg02_10g08"
/tissue_type="Mycelial tissue"
/dev_stage="Asexual"
/lab_host="E. coli (Sure cells)"
/clone_lib="Fg02_AAFCECORC_Fusarium graminearum mycelium"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: SmaI;
Site_2: XhoI; Mycelial tissue was collected from V8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (pL/UV) day light exposure. Mycelia was ground in

```

liquid nitrogen prior it's storage at -80 C until RNA extraction. Directional cloning with 5' end of cDNA cloned into EcoRI site of pBluescript and 3' end of cDNA cloned into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:
 Pred. No.: 1-26e-09 Length: 718
 Score: 199.50 Matches: 70
 Percent Similarity: 44.20% Conservative: 29
 Best Local Similarity: 31.2% Mismatches: 84
 Query Match: 8.82% Indels: 41
 DB: 12 Gaps: 9

US-09-985-689A-5 (1-433) x BI750157 (1-718)

QY 46 PheArgGlyLeuThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro 65
 Db 9 TTCGAGGSGTCGTCTCAGGCTGTCTACACTGCTTCAGCGGCCAGACGCT---GACACC 65
 QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLysGlyMet 85
 Db 66 AACGTCACGGAACCTCACCTGCTGCGCACTATGTCCGGAAGACATAC-----GGTGT 119
 QY 86 AlaProGlnAlaAsnLeu-----ValPheGln---SerIleVerAspSerSer 100
 Db 120 GCCAAGAGAGCCACCATCCAGCTGTCAAGGTCTTCCAGGGTAGTTTCATCCAGCACCTCC 179
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAnThrLeuPheSerGlnAlaTrpAsnAla 120
 Db 180 ATCATCTCTGCTGGCTTCACTGGGCTGCAACGACATCATCTCCAG----- 227
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn 140
 Db 228 ---GGCCGAACCAAGACATCGTCGTAATATGTCCTCGGGCGGTGTTACTCTGCTTCC 284
 QY 141 SerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaGly 160
 Db 285 TTCACAACGCTGTCAGTCTKCTCCAGCTCCGGTATTATCTCGCATGCTCCGGT 344
 QY 161 AsnGlyLeuProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 345 AACGATGGTGCCACAGCTCCACACTTCT---CCTGCTCTTCTCCAGCGCCACWACT 401
 QY 181 ValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsnHis 200
 Db 402 GTCGGTGCC-----ATTGACAGCAACTGGGCC 428
 QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
 Db 429 ATTKCTCTGACTCCAACTACGTTACCGTCTCTC-----GATATCTTT 470
 QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrp 240
 Db 471 GCVCCTKGCACCAAGCTTCTCTCGGC-----TGG 500
 QY 241 AlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
 Db 501 TACACGAGCAACAGTACCACCAACACCATCAGCGGACSTCCATGCTACTCTCCACATT 560
 QY 261 AlaGlyAsnVal 264
 Db 561 GCGGACTTGTG 572

RESULT 8

CA320325 771 bp mRNA linear EST 09-JUL-2003
 LOCUS UI-M-FWO-cby-d-23-0-UI.r1 NIH_BMAP_FWO Mus musculus cDNA clone
 DEFINITION IMAGE:6816072 5', mRNA sequence.
 ACCESSION CA320325
 VERSION CA320325.1 GI:24538449
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 711)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers

1..771
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGS:6816072"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_FWO"
 /note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction. Ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGGACAGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 3.62e-09 Length: 771
 Score: 195.50 Matches: 78
 Percent Similarity: 40.21% Conservative: 37
 Best Local Similarity: 27.27% Mismatches: 114
 Query Match: 8.65% Indels: 57
 DB: 14 Gaps: 10

US-09-985-689A-5 (1-433) x CA320325 (1-771)

QY 8 ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnValValAla 27
 Db 40 CTGCAGCAGATGCTGCTGG---CAGATGGGATACAGAGTCTTAATGTCAGAGTTGCT 96
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 97 GTTTTTCATCTGGCTC-----AGTGAGAAGCATCCGCATTTTAAG 138
 QY 48 GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
 Db 139 AAT-----GTGAGGAGAGAACCACTGGACCATGAGCGGACCGCTG 180
 QY 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
 Db 181 GATGATGGGTAGGCCATGGCACATCTGTTGCAAGTGTGATTCGACGATGAGGAGTGC 240

```

Qy 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly 101
Db 241 CAAGATTGCTCCAGATCGAGAGCTGCACATCTTTCAGGCTCTTTACCAACATCAG--- 297

Qy 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGly 121
Db 298 -----GTGCTTACACATCTTGGTTCTGTGATGCCCTTCACTATGCC 339

Qy 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSer 141
Db 340 ATCTAAAGAGATGAGGACCTTCTCACTTAGCATCGGTGGCCCGACCTTCATGGATCAT 399

Qy 142 ArgGlnValAspGluTyr-----ValArgAsnAspMetThrValLeuPheAlaAla 159
Db 400 CCGTTTGTTCAGAGGTGGGAATTAACAGCTAACATGTAATTAATGTTTCTGCTATT 459

Qy 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 179
Db 460 GGCATATGAGACCTCTCTATGGCACTCTGAATAACCTCTGATCATGATGGATGTGATT 519

Qy 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
Db 520 GGAGTGGGTGGC-----ATTGACTTTGAAGAT 546

Qy 200 HisIleAlaGlnPheSerArgGlyAlaThrArg-----AspGly 213
Db 547 AACATCGCTCGCTTTCTCCAGGGGAATGACTACCTGGGAATTACCGAGGAGCTATGGT 606

Qy 214 ArgIleIysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu 233
Db 607 CGTGTGAAGCCTGACATGTGACCTATGCTGCTGGAGTGGCGGGT----- 651

Qy 234 AlaProAspSerSerPheTrpAlaAsnTyrAsnSerIysTyrAlaTyrMetGlyGlyThr 253
Db 652 -----TCCGCTGTGAAGGGGGCTGCGGTGACCTCTCAGGGACC 690

Qy 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleIys 273
Db 691 AGTGTGCTTCCCGCAGTGGTGGTGGGGCGGTACCTTGTGTAGTANGCACAGTACAGAAG 750

Qy 274 AsnArgGlyIleThrPro 279
Db 751 CGGAGCTGGTGAATCCT 768

RESULT 9
BJ387574
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
Dictyostelium discoideum
VERSION
Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
SOURCE
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostelium discoideum at the slug stage
Unpublished (2002)
Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsinin@genes.nig.ac.jp.
Location/Qualifiers
1. 601
/mol_type="mRNA"
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dd3a18"

FEATURES
source

```

```

/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN

Alignment Scores:
Pred. No.: 3,78e-09 Length: 601
Score: 193.50 Matches: 55
Percent Similarity: 46.34% Conservative: 21
Best Local Similarity: 33.54% Mismatches: 49
Query Match: 8.56% Indels: 39
DB: 12 Gaps: 6

US-09-985-689A-5 (1-433) x BJ387574 (1-601)

Qy 197 AsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleIys 216
Db 51 AATGAGAATAATATTGTTTCATCTCAAAAGGTCCAAACACATGATGGTAGAATGAAA 110

Qy 217 ProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla----- 234
Db 111 CCTGATTTAGTTCCTGCTGGTGAATATATTACATCGGCAGATCAATGGTCCCAATACA 170

Qy 235 -----ProAspSerSerPheTrpAlaAsnTyrAsnSerIysTyrAlaTyrMet 250
Db 171 ACAGACCAATGTGGTGGTGGCTCTTTA---CCAATACAAATGCATTTATGGCG---ATA 224

Qy 251 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 270
Db 225 TCTGGTACATCAATGGCAACCTCATTCGACAGCAGCAACAAATCTTAGACAATAT 284

Qy 271 -----PheIleIysAsnArgGlyIleThrProIys 280
Db 285 TTAGTTGATGGTTATTATCCAACTGGTTCAATTTAGAAATCAAAATAAATTACCACT 344

Qy 281 ProSerLeuIleIysAlaAlaLeuIleAlaGlyAla----- 292
Db 345 GGATCATTTATTAAGCATTAATGATTAATATCTCAGTTATTAATGGTACATTCAA 404

Qy 293 -----ThrAspValGlyLeuGlyTyrProSerGlyAsp----- 303
Db 405 TTGATTACATCATCAAGTATTACATATCCATCAACCAAGTTTGTGAAATTTTGCAGGT 464

Qy 304 -----GlnGlyTrpGlyArgValThrLeuAspIysSerLeuAsnValAlaTyr 319
Db 465 GCAAGTTTAGTTCAGGTGGGTGGTCTATTAGAAATGAGTAATGGTTACATGTTGTCAT 524

Qy 320 ValAsnGluAla 323
Db 525 AATAATAATACT 536

RESULT 10
BJ387574
LOCUS
DEFINITION
Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION
Dictyostelium discoideum
VERSION
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
SOURCE
Toxoplasma gondii
ORGANISM
Toxoplasma gondii
REFERENCE
Tang,K., Cole,R., Fogarty,S., Sibley,L.D., Ajioke,J.A., White,M.,
AUTHORS
Clifton,S., Pape,D., Martin,J., Wylie,T., Dente,M., Warra,M.,
Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M.,
Ritter,S., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I.,
Kennedy,S., Maguire,L., Waterston,R. and Wilson,R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.

```


Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxos@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 445.
Location/Qualifiers
1. 545
/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="VGS (Type III)"
/db_xref="taxon:5811"
/clone="TGSTzy80b07.y1"
/dev_stage="partially sporulated oocysts"
/note="Vector: Modified pBluescript (pBSSK+); Site 1:
BamHI; Site 2: EcoRI; PolyA mRNA from partially-sporulated
oocysts was converted to cDNA using the
template-switching PCR method (SMART cDNA, Clontech Inc.,)
and sized selected on SizeSep 400 columns (Amersham
Pharmacia Biotech Inc.). First strand was reverse
transcribed using the CDS III-oligo-dT primer and a 5'
template switch primer (Smart IV primer). The product of
the first strand synthesis was PCR amplified using the
same primer set and the fragments were digested with SfiI.
The fragments were size selected, ligated into a modified
pBluescript vector containing directional SfiI sites, and
electroporated into DH10B or DH12S cells. Vector: SfiI
sites were added to the multiple cloning region of
pBluescript SK+ between the BamHI/EcoRI sites. The
modified polylinker has the following sequence:
5'GATTCGGCATACGGCC(G)n- insert--
GGCGCTCGGCGCCGAGACCGC3' where n=3-4 G nucleotides. Library
Source: Michael White, Maria E. Jerome, Emily A. Johnson,
Jay A. Radke, Montana State University. Clone
Availability: David Sibley, Washington University"

FEATURES

source

Alignment Scores:
Pred. No.: 7,28e-09 Length: 545
Score: 190.00 Matches: 64
Percent Similarity: 44.81% Conservative: 31
Best Local Similarity: 30.19% Mismatches: 75
Query Match: 8.40% Indels: 42
DB: 13 Gaps: 9

US-09-985-689a-5 (1-433) x BU575479 (1-545)

QY 64 AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
Db 4 GACGACACGGCCACGGCAGCGATGTCCGGC---ATCATCGGGCGATCAAGGACCAA 60
QY 79 AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
Db 61 AACTCCCGGTGAAGAGCGCTCTGTGGAAACACAGCATTCGCCGCTTAAGTTCATGGGT 120
QY 99 SerSerGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db 121 GCAACGGA-----AACGGTCGACGTCAGATGCCATCAAGGCTCTCAACTACGCGGTC 174
QY 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrp 138
Db 175 CAGATCGGAATTCCTCGTCACTGCTGGCGCGGTCCACCTGTCGAGCGGTG 234
QY 139 AlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAla 158
Db 235 ATTGCGGCTCTGAACGGCCGAGCGTGGTCAC-----CTTTTCATTCGCGG 285
QY 159 AlaGlyAsnGluGlyProAsnSerGlyThrIle-----SerAlaProGlyThrAla 175

Db 286 GCGGGAACCAAGGAAGAACCCGACGAATTCGCATTACCTCCCTCGTACCGGCTC 345
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAla 195
Db 346 GCAACCTAGTCAGCGTGGCTGCACA----- 372
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 373 AACTCCGAGGACCACTCGCTCCTTCAGAACCGGGAGCGGCC----- 417
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db 418 ACGGTTGACCTTCGGCGCCCGCGGTAAATCCTC-----TCTACCTTTCGCCCA 468
QY 236 AspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMet 255
Db 469 GAC-----CAGTTTCGGGAACCTTTCGGGAACCTCCATG 501
QY 256 AlaThrProIleValAlaGlyAsnValAlaGlnLeu 267
Db 502 GCTACGCGCGTGTCTGGGGCGTCGCGCGCATCCTC 537

RESULT 11
TA315H10P
LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 315h10, forward sequence,
genomic survey sequence.
ACCESSION
AL490202
VERSION
AL490202.1 GI:11866292
KEYWORDS
GSS.
ORGANISM
Trypanosoma brucei
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
1 (bases 1 to 574)
Hall, N., Bowman, S., Lennard, N. J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
McNeill, S. E., Rajadream, M. A. and Barrell, B. G.
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
COMMENT
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU27/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J. C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nh@sanger.ac.uk
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES
source
1. 574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU27"
/db_xref="taxon:5691"
/clone="315h10"

ORIGIN
Alignment Scores:
Pred. No.: 1.27e-08 Length: 574
Score: 188.00 Matches: 51
Percent Similarity: 47.65% Conservative: 30
Best Local Similarity: 30.00% Mismatches: 59
Query Match: 8.31% Indels: 30
DB: 29 Gaps: 6

US-09-985-689A-5 (1-433) x TA315H10P (1-574)	
QY	18 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
Db	63 GGTATTGACGGTAGTGGTACAGTAATAGGGGTAGCGGATACGGTATCGACTTT-----116
QY	38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db	117 AACAGCTGTTTCTTCACGATCCA-----AATCAAGAGGTGCGCTTTACCCGAAGGTT 170
QY	58 ArgThrAsnAsn-----61
Db	171 AACTATACACCGCAAAATCGTGTCTATTGCCCCGTGTGACTTCATCCGGGGGATTAC 230
QY	62 -----AlaAsnAspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
Db	231 TTTGCTGGGATGAGGAAATAGTTCATGCGACGACGACGTCGCGAGGTACCCGACGGGGAGT 290
QY	76 ValLeu-----GlyAsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db	291 GTTATTAGTAAACGACGGTAATGCAAGTAGTATATGTTGTGCCAAGGGGGCGAAGTTTC 350
QY	93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThr 112
Db	351 TTCAGGGGTTGGTGGCCATCCAGTCAGAGCTTGTCTCCGCCACGACGTACTCAA 410
QY	113 LeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 132
Db	411 ATTATTCGTCGGATATGCGCTGGAGCCCGTGTCTCAAACTCGTGGGTTTGT 470
QY	133 ValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsn---Asn 151
Db	471 GCTCCCTCCGATATTCTGCTGTGGAAAGGATATGATGAGTTTGGAGTAGTTATGAC 530
QY	152 AspMetThrValLeuPheAlaAlaGlyAsn 161
Db	531 GATCGCTACTTATCTTCCACATGGCAAC 560
RESULT 12	
AJ273402 614 bp mRNA linear EST 29-DEC-1999	
LOCUS AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae	
DEFINITION CNA clone Ma#948, mRNA sequence.	
ACCESSION AJ273402.1 GI:6432774	
VERSION EST.	
KEYWORDS Metarhizium anisopliae	
SOURCE Metarhizium anisopliae	
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; Clavicipitaceae; Metarhizium.	
REFERENCE 1 (bases 1 to 614)	
AUTHORS Screen,S.E., Mathur,P. and St. Leger,R.J.	
TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae	
JOURNAL Unpublished (1999)	
COMMENT Contact: Screen SE	
Entomology	
University of Maryland	
4112 Plant Sciences Building, College Park, MD 20742, USA.	
Location/Qualifiers	
1. .614	
/organism="Metarhizium anisopliae"	
/mol_type="mRNA"	
/strain="ARSEF 2575"	
/db_xref="taxon:5530"	
/clone="Ma#948"	
/clone_lib="Metarhizium anisopliae ARSEF 2575"	
/note="Vector: UniZap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector, UniZap"	
ORIGIN	
Alignment Scores:	
US-09-985-689A-5 (1-433) x AJ273402 (1-614)	
QY	21 GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db	11 GGTGAGGTACTTCCGTATATATCATGTACACTGGTATTGAG-----52
QY	41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db	53 GCCTCCACCCCGAGTTTGGGTCGGCCACTTTTCTTAAGAGCTTCATCAGCGGTCAA 112
QY	61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db	113 AAC---ACTGATGCCACGGCCATGGGACTCACTGGCTGTGATCACTT-----GGTAGC 163
QY	81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer--- 99
Db	164 AAGACCTACGCTGTTCGCAAAAGGCTTAAGCTCTATGCTGTCAGGTTCTTGACACACG 223
QY	100 ---SerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
Db	224 GGCAGTGGTCTCTACTCCGCTATCATCATGGTGGCATGACTACGTTCACAGGCTCCAAG 283
QY	119 AsnAlaGly-----AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGly 135
Db	284 ACCCGCGCTGCCCAACGGCGCATTTGCTTCATGAGCTGGGA-----GGT 331
QY	136 AlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnMetThrVal 155
Db	332 GGCCTACTCGGCTCCGTCACCAACCAAGGTGCTGCTTTGGTCAATTCGTGTCTTCTT 391
QY	156 LeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAla 175
Db	392 CCGCTGCGCGTGGCAAGGATACCGGATGCCAGAACACCTCT---CCGCTTCGAG 448
QY	176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAla 195
Db	449 CTTCTGCTGCTGCTGTTGGTGGCTCT-----GCG 478
QY	196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db	479 GAAATGACAGCGCGATCTTCTTCTTCAACTAC-----GGCAGAGTT 520
QY	216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
Db	521 ---GTCGATATTTGCTCTCTGTTAGCAATGTTCTTCCACC-----559
QY	236 AspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMet 255
Db	560 -----TGG-----ATTGGTGGCGGCACAAACACCATCTCTGTGTACCTNCATG 601
QY	256 AlaThrPro 258
Db	602 GCTACTTCCC 610
RESULT 13	
BZ424995/c 2121 bp DNA linear GSS 13-DEC-2002	
LOCUS BZ424995 Aspergillus terreus random genomic DNA clone library	
DEFINITION Aspergillus terreus genomic, genomic survey sequence.	
ACCESSION BZ424995	
VERSION BZ424995.1 GI:26666450	
KEYWORDS GSS.	
SOURCE Aspergillus terreus	
ORGANISM Aspergillus terreus	
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.	

```

REFERENCE
AUTHORS
1 (bases 1 to 2121)
Askenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S.,
Zimmer,D.P., Boers,M.E., Blomquist,P.R., Martinez,E.J.,
Monreal,A.W., Feibelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K.,
Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and
Madden,K.T.
TITLE
Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
JOURNAL
Unpublished (2002)
COMMENT
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
FEATURES
Location/Qualifiers
1..2121
/organism="Aspergillus terreus"
/mol_type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db_xref="taxon:31178"
/lab_host="Escherichia coli"
/clone_lib="Aspergillus terreus random genomic DNA clone
library"
/note="Vector: pZEROTM-2; Site 1: Sau3A; Site 2: BamHI;
Sau3A genomic fragments ligated into BamHI digested
pZEROTM-2"
ORIGIN
Alignment Scores:
Pred. No.: 3.4e-07 Length: 2121
Score: 183.50 Matches: 79
Percent Similarity: 38.97% Conservative: 34
Best Local Similarity: 27.24% Mismatches: 86
Query Match: 8.12% Indels: 92
DB: 28 Gaps: 12
US-09-985-689A-5 (1-433) x BZ424995 (1-2121)
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1281 GGGGAGGGGACCTACCTTACGTCGTAGACACGGGACATCAAGTGGAC-----1234
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
Db 1233 -----CACGAGGAATTTGAGGGCCGTCGGAGTCTCGCTACACGCTCCGGTGGCCAG 1180
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1179 CATGTG---GACGGTGTGGCCATGGACCCACCGTTTCTGTACATTGGTGTAGACA 1123
QY 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle-MetAspSerSe 100
Db 1122 TAC-----GGCGTGGCCAGAGAGGCCACTGCTG-----TCGGTCAAGGTCTTTCGTC 1075
QY 100 rGlyGlyLeuGlyGlyLeuProSerAsnLeuAsn-----ThrLe 113
Db 1074 GGGGAATCGAGCAGCACATCATCTCTGGACGGCTTCACCTGGGGCCCATGACATT 1015
QY 113 uPheSerGln-----AlaTrpAsnAlaGlyAlaAr 123
Db 1014 GTTTCGAAGAGCGCACTGGAAAGGCGAGCATCAACATGAGCTTGGGTATGTTATTCTGC 955
QY 123 GileHisThrAsnSerTrp-----GlyAlaProValAsnGlyValaty 137
Db 954 CATCTATGTGAATGA-TATCATTTGGAGTCCACAGCTGACGAGCATGATGGCGGTGGATA 996
QY 137 rThrAlaAsnSerArgGlnValaAspGluTyrValArgAsnAsnAspMetThrValLeuPh 157
Db 895 CTCGAAGCCCTCAATGATCGCGTGGAGAGCGCTTTTCACGAGGAGTCTGTCCTCGT 836
QY 157 eAlaAlaGlyAsnGluGly-----ProAsnSerGlyThrIleSer-- 170

```

```

Db 835 CGCCGCGGCATGAGATGTGTCGTCACCGTTCACCCCTGTGAATGGAGTCCGACTCC 776
QY 171 -----AlaProGly-ThrAlaLysAsnAlaI 179
Db 775 GCTGACTGATTCGACAGACCGAGCGCTGAGCAGCCCGCTTCTGCTCTGATGCC 716
QY 179 leThrValGlyAla-----ThrGluAsnTyArgProSerPheGlySerI 194
Db 715 TTACCGTCTGCTGGATCAACGTAACACACCGCTGCTATTCTTCCAACTACGGGTCCG 656
QY 194 euAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrAgApGlyA 214
Db 655 TGGTG----- 651
QY 214 rGileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuA 234
Db 650 -----GATATCTTCGCCCCGGCCAGACATCTCTCTGCC----- 615
QY 234 laProAspSerSerPheTrpAlaAsnTyArgSerLysTyArgAlaTyMetGlyGlyThrs 254
Db 614 -----TGGATCGGCTCCACACGCGCCACCAACACCATCTCGGGCACCT 572
QY 254 erYetaAlaThrProIleValAlaGly 262
Db 571 CCATGGCCACCCCACTGTCTGCC 546
RESULT 14
LOCUS BQ770462 665 bp mRNA linear EST 26-JUL-2002
DEFINITION UI-M-PIO-byV-m-19-0-UI.r1 NIH BMAP_F10 Mus musculus cDNA clone
IMAG:5702970 5', mRNA sequence.
ACCESSION BQ770462 GI:21978936
VERSION BQ770462.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
NH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgahe-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..665
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAG:5702970"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_F10"
/note="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated

```

with EcoR I adaptor, digested with NotI and then cloned directionally into pVX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CACCACGAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): "Gene discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:
 Pred. No.: 5.93e-08 Length: 665
 Score: 182.50 Matches: 65
 Percent Similarity: 42.68% Conservative: 37
 Best Local Similarity: 27.20% Mismatches: 96
 Query Match: 8.07% Indels: 41
 DB: 13 Gaps: 7

US-09-985-689A-5 (1-433) x BQ770462 (1-665)

```

QY 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn 82
DB 14 GATGATGGGTAGGCCATGCGACATTCGTTGAGGTGTGATTCGCCAGCATGAGGGAGTGC 73
QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
DB 74 CAGGATTGCTCCAGATGACAGATCCATCTTCAGGGCTTTTACCAACATCAG--- 130
QY 102 GlyLeuGlyGlyLeuProSerAsnLeuAsnThrPheSerGlnAlaThrAlaProAlaGly 121
DB 131 -----GTGCTTACACATCTTGTGTTCTGATGCGCTTCAACTATGCC 172
QY 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaThrAlaAsnSer 141
DB 173 ATCTTAAGAGATGAGACCTTCAACCTTAGCATCGGTGGCGCGGACTTCATGAGTATC 232
QY 142 ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla 159
DB 233 CCGTTTGTTCACAGGTGTGGGATTAAACAGCTAACATGTAATTATGTTTCTGCTATT 292
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
DB 293 GCGAATGATGACCTCTCTATGGACACTCTGAATTAACCTCTGATCAGATGGATGTGATT 352
QY 180 ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerLeuAlaAspAsnProAsn 199
DB 353 CGAGTGGGTGCG-----ATTGACTTTGAGAT 379
QY 200 HisIleAlaGlnPheSerSerArgGlyAlaThrArg-----AspGly 213
DB 380 AACATCGCTCGCTTTCTTCCAGGGGAATGACTACCTGGGAATTACCAGGAGGCTATGGT 439
QY 214 ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu 233
DB 440 CQTGTGAAGCTTGACATTTGCTACCTATGCTGCTGGAGTGGGGT----- 484
QY 234 AlaProAspSerPheThrAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr 253
DB 485 -----TCCGGTGTGAAGGGGGCTCCCGTGCACTCTCAGGGACC 523
QY 254 SerMetAlaThrProIleValAlaGlnValAlaGlnLeuArgGluHisPheIleLys 273
DB 524 AGTGTGCTTCNCCAGTGTGCTGGGCGGCTACCTTTGTTAGTAAAGCAGATACAGAG 583
QY 274 AsnArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAla 292
DB 584 CGGGAGCTGGTG-----AATCCTCCAGTGTGAAGCAGCGCTTTGATAGCTGACGCC 634

```

RESULT 15

BZ893395/c
 LOCUS BZ893395 716 bp DNA linear GSS 30-JUL-2003
 DEFINITION HL2.0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
 genomic survey sequence.

BZ893395
 BZ893395.1 GI:33343985
 GSS.
 Halorubrum lacusprofundi
 Halorubrum lacusprofundi
 Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 Halobacteriaceae; Halorubrum.
 1 (Bases 1 to 716)
 Goo, Y., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M.,
 DasSarma, S., Ng, W.V. and Hood, L.
 Low-pass Sequencing for Microbial Comparative Genomics
 Unpublished (2003)
 Contact: Goo Y
 Institute for Systems Biology
 1441 North 34th Street, Seattle, WA 98103, USA
 Tel: 206 732 1412
 Fax: 206 732 1299
 Email: ygoosystemsbiology.org
 Seq primer: M13 Forward
 Class: shotgun.
 Location/Qualifiers
 1..716
 /organism="Halorubrum lacusprofundi"
 /mol_type="genomic DNA"
 /strain="ATCC 49239"
 /db_xref="taxon:2247"
 /clone_lib="H1 pUC18 Library"
 /note="Vector: pUC18, site 1: SmaI; A shotgun library was
 constructed from Halorubrum lacusprofundi genomic DNA
 using pUC18/SmaI/BAP plasmid"

ORIGIN

Alignment Scores:
 Pred. No.: 6.73e-08 Length: 716
 Score: 182.50 Matches: 71
 Percent Similarity: 45.45% Conservative: 34
 Best Local Similarity: 30.74% Mismatches: 66
 Query Match: 8.07% Indels: 61
 DB: 28 Gaps: 11

US-09-985-689A-5 (1-433) x BZ893395 (1-716)

```

QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu---Gly 78
DB 701 TCCGATGCTCGACGACGAGCGGACGAACTCAGTCGCGGACCGTTCGCGGGCGGA 642
QY 79 AsnAlaLeuAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 96
DB 641 AACGCGAGTGGGACCGCAATCGGCTGCGACCGAACGCGTCTCCACGGGATCAAGGTG 582
QY 97 MetAsp-SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGly 116
DB 581 TTCGACGACGACGGGA-----CCAAACGGACGTTCTGTCGGGTCTGTCGGC 537
QY 116 nAlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrp-----GlyAlaProVa 133
DB 536 GGAATGGACACGCGACGCGAGGATCGGACGTCGACGTCGACATCAGCTCGGTGCG 477
QY 133 LAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAspMe 153
DB 476 GACGGGCACTTACCATCTTCTATCATCAACCG-GTT---CGGAACACTCGAGTCCGGGAA 421
QY 153 tThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGly 173
DB 420 GATCGCGCTGTTTTCGCGCGGAGACATCGGT-----CAAGGAACGTCGAGTCTCCCGG 367
QY 173 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySe 193
DB 366 GAACGCTACGACTCGCTCGCGTCGGAGCGGTCAACGACGCCG----- 321
QY 193 rLeuAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThr----- 210
DB 320 -----GGCGTCCGCCCACTTCTCCAGC---GGAGACACCATCAACAC 283

```

```

QY 211 -----ArgAspGlyArgIleLysProAs 218
Db 282 GTCAGCGCGTGGGAAGTATGCCCGCGGACTGGCCCGACGAGTACGTGGTTCGGGA 223
QY 218 pValThrAlaProGlyThrPheIleLeuSerAlaArg-----SerSerLeuAlaPr 235
Db 222 CGTGTGGCGCGCAGCGGTGAGCGGTATCTCGCGGACCGGGCGGTACAAACATCCGAAA 163
QY 235 oAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSerMe 255
Db 162 GGAC-----GGCACCTCCAT 148
QY 255 tAlaThrProIleValAlaGlyAsnValAla 265
Db 147 GGCCGACCGCACGTCAGCGCGTCCGGCG 117

```

Search completed: April 4, 2004, 11:55:04
Job time : 2283.58 secs

/strain="VEG (Type III)"
 /db_xref="taxon:5811"
 /clone="IGSRzy80807.y1"
 /dev_stage="partially sporulated oocysts"
 /clone_lib="79VEG partially sporulated oocyst cDNA"
 /notes="vector: Modified pBluescript (pBSK+); Site 1:
 BamHI; Site 2: EcoRI; PolyA mRNA from partially-sporulated
 oocysts was converted to cDNA using the
 template-switching PCR method (SMART cDNA, Clontech Inc.)
 and sized selected on SizeSep 400 columns (Amersham
 Pharmacia Biotech Inc.). First strand was reverse
 transcribed using the CDS III-oligo-dT primer and a 5'
 template switch primer (Smart IV primer). The product of
 the first strand synthesis was PCR amplified using the
 same primer set and the fragments were digested with SfiI.
 The fragments were size selected, ligated into a modified
 pBluescript vector containing directional SfiI sites, and
 electroporated into DH10B or DH12S cells. Vector: SfiI
 sites were added to the multiple cloning region of
 pBluescript SK+ between the BamHI/EcoRI sites. The
 modified polylinker has the following sequence:
 5'GAATTCGGCGCATTCAGGCG(G)n-- insert--
 GGCGCTCGCGCCACGATCC3'where n=3-4 G nucleotides. Library
 Source: Michael White, Maria E. Jerome, Emily A. Johnson,
 Jay A. Radke, Montana State University. Clone
 Availability: David Sibley, Washington University"

ORIGIN

Alignment Scores:
 Pred. No.: 5,24e-09 Length: 545
 Score: 190.00 Matches: 64
 Percent Similarity: 44.81% Conservativity: 31
 Best Local Similarity: 30.19% Mismatches: 75
 Query Match: 8.41% Indels: 42
 DB: 13 Gaps: 9

US-09-985-689A-4 (1-433) x BUS75479 (1-545)

Qy 64 AspproAnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
 Db 4 GACGACCCAGCGCCAGCGCATGTCGCCGCGC---ATCATCGGGGGCATCAGGACCAA 60
 Qy 79 AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
 Db 61 AACTCCCGGGTGMAGGGGTCTGTGGGAACACGAGCATTCGCCGCTTAAGTTTCATGGGT 120
 Qy 99 SerSerGlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
 Db 121 GCAAAACGGA-----AACGGGTGACGTCAGATGCCATCAGAGCTCTCACTACGCGGTC 174
 Qy 119 AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThr 138
 Db 175 CAGATGGGAATTCCTCAGCTGCACTGCTGGGCGGTCCACCTGGTCCGAGCGCTG 234
 Qy 139 AlaAsnSerArgGlnValAspGluTrpValArgAsnAsnAspMetThrValLeuPheAla 158
 Db 235 ATTGGGGTCTGTGGAAGCGCGAGAGCGTCGGTCAC-----CTTTTCATTGGCGCG 285
 Qy 159 AlaGlyAsnGluGlyProAsnSerGlyThrIle-----SerAlaProGlyThrAla 175
 Db 286 GCGGGAACCAAGAGAGAACACCGACGAATTCGCCATTACCTCGCTCCGCGGTC 345
 Qy 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTrpArgProSerPheGlySerIleAla 195
 Db 346 GCAAAACGTAGTCAGCGTGGCTCGACA----- 372
 Qy 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
 Db 373 AACTCGGAGGACCACTCGCTCCCTTCACCAACCGGGGAGCGGCC----- 417
 Qy 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 235
 Db 418 ACGGTTGACCTTGGCGCGCGCGGTAAATCTCTC-----TCTACCTTTCGCCCA 468

Qy 236 AspSerSerPheTrpAlaAsnTrpAsnSerLysTrpAlaTrpMetGlyGlyThrSerMet 255
 Db 469 GAC-----CAGTTTCGCAACTTTCGCGAACCTCCATG 501
 Qy 256 AlaThrProLeuValAlaGlyAsnValAlaGlnLeu 267
 Db 502 GCTAGCGCGGTGGTGGCGGGCGTCCGCCCATCCTC 537

RESULT 12

AJ273402 614 bp tRNA linear EST 29-DEC-1999
 LOCUS AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
 DEFINITION CDNA clone Ma#948, mRNA sequence.

ACCESSION AJ273402
 VERSION AJ273402.1 GI:6432774
 KEYWORDS EST.
 SOURCE Metarhizium anisopliae
 ORGANISM Metarhizium anisopliae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocryomycetidae; Hypocreales; Clavicipitaceae; Mitosporic
 Clavicipitaceae; Metarhizium.

REFERENCE 1 (bases 1 to 614)
 AUTHORS Screen,S.E., Mathur,P. and St. Leger,R.J.
 TITLE EST analysis of the insect pathogenic fungus Metarhizium anisopliae
 JOURNAL Unpublished (1999)
 COMMENT Contact: Screen SE
 Entomology

University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA.

FEATURES

Location/Qualifiers
 1..614
 /organism="Metarhizium anisopliae"
 /mol_type="mRNA"
 /strain="ARSEF 2575"
 /db_xref="taxon:5530"
 /clone="Ma#948"
 /clone_lib="Metarhizium anisopliae ARSEF 2575"
 /note="Vector: UniZap; Metarhizium anisopliae was grown on
 insect cuticle for 24 hours. A cDNA library was
 constructed in the unidirectional Lambda vector, UniZap"

ORIGIN

Alignment Scores:
 Pred. No.: 8,14e-09 Length: 614
 Score: 189.00 Matches: 68
 Percent Similarity: 46.09% Conservativity: 44
 Best Local Similarity: 27.98% Mismatches: 83
 Query Match: 8.36% Indels: 48
 DB: 9 Gaps: 12

US-09-985-689A-4 (1-433) x AJ273402 (1-614)

Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 11 GGTCAGGGTACTGCGTATATATCATTCAGACCTGGTATTGAG----- 52
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTrpAlaLeuGlyArgThrAsn 60
 Db 53 GCCTCCACCCCGAGTTTGAGGGTCGCGCCACTTTTCTTAAGAGCTTCATCAGCGGTCAA 112
 Qy 61 AsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 Db 113 AAC---ACTGATGGCCACCGCCATGGGACTCACTGCGCTGGTACCATT-----GGTAGC 163
 Qy 81 LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer--- 99
 Db 164 AAGACCTACCGTGTGCCAAAAGGCTATGCTGTCAAGTGTCTTCAAGTTCTTGACCAACAG 223
 Qy 100 ---SerGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp 118
 Db 224 GGCAGTGGTTCTACTCCGGGTATCATCAGTGGCATGGACTACGTTGCACAGGACTCCAG 293

```

Db      790 TGGAGTGGCGACTCCGCTGACTGATTGACAGACCGCGCTGGAGCACCGCCGGCTT 731
Qy      174 hrAlaLysAenAlaIleThrValGlyAla
Db      730 CTGCTCTGATGCGCTTACGCTCGCTGCGATCAACGTAACACCGCTGCTATTCT 671
Qy      189 roSerPheGlySerIleAlaAspAenProAsnHisIleAlaGlnPheSerSerArgGlyA 209
Db      670 CCAACTACGCTCGCTGGTG-
Qy      209 laThrArgAspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerA 229
Db      650 -----GATATCTTCGCGCGCGGCGGCGAGACATCTCTCTG 617
Qy      229 laArgSerSerLeuAlaProAspSerPheTrpAlaAenTyraenSerLysTyraIar 249
Db      616 CC-----TGGATCGCTCTCAACACGCGCCACCAACA 587
Qy      249 yrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAenValAlaGlnLeuArg- 268
Db      586 CCAATCGGCACTCATGCGCCACCCCAATGTGCGCTGCTCCATCTACTTGATGT 527
Qy      269 -----GluHisPheIleLysAenArgGlyIleThrProLysProSerLeuIleLysAlaA 287
Db      526 CGTGGAGGACCTCAGCAGCCCAAGCGCTCACTGACCGC-----ATCAAGGAGC 476
Qy      287 laLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyroProan 301
Db      475 TAGCGACTGAGCGCTGCTCAGCAACGTTGCGCGTAGCCCCAAC 432

```

```

RESULT 10
LOCUS   BJ387574
DEFINITION BJ387574 Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION BJ387574
VERSION   BJ387574.1 GI:19296958
KEYWORDS EST.
SOURCE   Dictyostelium discoideum
ORGANISM Dictyostelium discoideum

```

```

REFERENCE 1 (bases 1 to 601)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

```

```

FEATURES
source 1..601
         /organism="Dictyostelium discoideum"
         /mol_type="mRNA"
         /strain="AX4"
         /db_xref="taxon:44689"
         /clone="dds3a18"
         /sex="mat A"
         /dev_stage="slug stage"
         /clone_lib="Dictyostelium discoideum cDNA library, SF"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 4.34e-09
Score: 191.50
Length: 601
Matches: 54
Conservative: 23
Best Local Similarity: 46.95%
Mismatches: 48
Query Match: 32.93%
Indels: 39
Gaps: 6
DB: 12

```

US-09-985-689a-4 (1-433) x BJ387574 (1-601)

```

Qy      197 AsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLys 216
Db      51 AATCAGATATATATTTGTTTCATTTCTCATCAAAAGGTCACACACATGATGGTAGAATGAAA 110
Qy      217 ProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla----- 234
Db      111 CCTGATTAGTTGGCTTGGTGAATATATATCATCGGCAAGATCAAAATGGTGGCAATACA 170
Qy      235 -----ProAspSerSerPheTrpAlaAsnTyraenSerLysTyraIaTyMet 250
Db      171 ACAGACCAATGGTGGTGGCTCTTTA---CCAAATACAAATGCATTTATGGCG---ATA 224
Qy      251 GlyGlyThrSerMetAlaThrProIleValAlaGlyAenValAlaGlnLeuArgGluHis 270
Db      225 TCTGTTACATCAATGGCAACCTCATTTTGCAGCAGCAGCAACAACAATCTTTAGACAATAT 284
Qy      271 -----PheIleLysAenArgGlyIleThrProLys 280
Db      285 TTAGTTGATGTTTATTCCAACTGGTTCATTTAGTAATCAATAATTAATACACCAACT 344
Qy      281 ProSerLeuIleLysAlaLeuIleAlaGlyAla----- 292
Db      345 GGATCATTTATTAAGCATTAATGATTAATAATGCTCAGTTTATTAATGGTACATTTCAA 404
Qy      293 -----ThrAspValGlyLeuGlyTyroProAsnGlyAsp----- 303
Db      405 TTGATTACATCATCAAGTATTCATATCCATCAACCAACCAAGTTTTTGAATAATTTGCAGGT 464
Qy      304 -----GlnGlyTrpGlyValThrLeuAenLysSerLeuAenValAlaTy 319
Db      465 GCAAGTTTGTAGTTCAAGGTTGGGTGCTATTAGTAATGAGTAATGGTTACATGTTGTCAT 524
Qy      320 ValAenGluAla 323
Db      525 AATAATAATAGT 536

```

RESULT 11

LOCUS BJ575479

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..545

/organism="Toxoplasma gondii"

/mol_type="mRNA"

EST. 17-SEP-2002

TGSTRYB80B07.Y1 TGVEG partially sporulated oocyst cDNA Toxoplasma

Gondii cDNA clone TGSTRYB80B07.Y1 5' similar to TR.Q45522 Q45522

PREPRO-SUBSTITUTIN SENDAI PRECURSOR. ; mRNA sequence.

BU575479

BU575479.1 GI:23045230

EST.

Toxoplasma gondii

Toxoplasma gondii

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;

Sarcocystidae; Toxoplasma.

1 (bases 1 to 545)

Tang,K., Cole,R., Fogarty,S., Sibley,L.D., Ajoka,J.A., White,M.,

Clifton,S., Pape,D., Martin,J., Wyllie,T., Dante,M., Marra,M.,

Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M.,

Ritter,E., Bennett,J., Franklin,C., Tsagaris,I., Ronko,I.,

Kennedy,S., Maguire,L., Waterston,R. and Wilson,R.

Toxoplasma EST Project

Unpublished (2001)

Contact: Clifton, S.

Toxoplasma EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: toxo@wustl.edu

Contact David Sibley (toxest@borcim.wustl.edu) for further

information relating to organism, libraries, or clone availability.

Seg primer: -40RP from Gibco

High quality sequence stop: 445.

Location/Qualifiers

1..545

/organism="Toxoplasma gondii"

/mol_type="mRNA"

Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Db	97	GTITTTGATACTAGGGCTC-----AGTGAGAAGCATCGCATTTTAAAG	138
Qy	48	GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn-----AsnAla	62
Db	139	AAT-----GTGAGGAGAGAACCACACTGGACCATGAGCGGACCCTG	180
Qy	63	SerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn	82
Db	181	GATGATGGGCTAGGCCATGGCACATTCGTTCCAGGTGTGATTGCCAGCATGAGGGAGTCC	240
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly	101
Db	241	CAAGGATTTGGCTCCAGATGCGAGCTGCACATCTCAGGGTCTTTACCAACATCAG---	297
Qy	102	GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTyrPheAlaGly	121
Db	298	-----GTGCTTTACACATCTTGGTTTCTGGATGCCCTCACTATGCC	339
Qy	122	AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsnSer	141
Db	340	ATCCTAAGAAGATGGAGGTTCTCAACCTTAGCATCGTGGGCCCGCACTTCATGGATCAT	399
Qy	142	ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla	159
Db	400	CGGTTTGTGGACAAGGTGGGAATTAACACTAACCAATGTAAATTATGGTTTCTGCTATT	459
Qy	160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179
Db	460	GGCAATGATGGACCTCTCTATGGCACTCTGAATAACCTCTGATCAGATGGATGTATT	519
Qy	180	ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn	199
Db	520	GGAGTGGGTGGC-----ATTGACTTTGAAGAT	546
Qy	200	HisIleAlaGlnPheSerSerArgGlyValaThrArg-----AspGly	213
Db	547	AACATCGCTCGCTTTCTTCCAGGGAATGACTACCTGGGAATTACACAGAGGCTATGT	606
Qy	214	ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu	233
Db	607	CGTGTCAAGCCTGCATGCTACCTATGGTCTGGAGTGGGGGT-----	651
Qy	234	AlaProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr	253
Db	652	-----TCCGGTGTGAAAGGGGGGTGCGGTGCACCTCTCAGGGACC	690
Qy	254	SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuLeuArgGluHisPheIleLys	273
Db	691	AGTGTGCTCTCCAGTGTGTGCTGGGGGGGTACCTTGTGTAGTANGCACAGTACAGAG	750
Qy	274	AsnArgGlyIleThrPro	279
Db	751	CGGGAGCTGTGTAATCCT	768

RESULT	9
BZ424995/c	
LOCUS	BZ424995.1
DEFINITION	Aspergillus terreus genomic DNA clone library
ACCESSION	BZ424995
VERSION	BZ424995.1
KEYWORDS	GSS.
SOURCE	Aspergillus terreus
ORGANISM	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
REFERENCE	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
AUTHORS	Askenazi,M., Driggers,E.M., Holtzman,D.A., Norman,T.C., Iverson,S., Zimmer,D.P., Boers,N-E., Blouquist,P.R., Martinez,E.J., Monreal,A.W., Feilbelman,T.P., Mayorga,M.E., Maxon,M.E., Sykes,K., Tobin,J., Cordero,E., Salama,S.R., Trueheart,J., Royer,J.C. and 1 (bases 1 to 2121)

```

Madden, K.T.
Integrating transcriptional and metabolite profiles to direct the
engineering of lovastatin-producing strains
JOURNAL
Unpublished (2002)
COMMENT
Contact: Zimmer DP
Microbia, Inc.
One Kendall Square Building 1400 W, Cambridge, MA 02139, USA
Tel: 617-621-8322
Fax: 617-
Email: dzimmer@microbia.com
Class: plasmid ends.
Location/Qualifiers
1. .2121
/organism="Aspergillus terreus"
/mol type="genomic DNA"
/strain="ATCC 20542 (A. terreus Thom, anamorph)"
/db xref="taxon:33178"
/lab hosts="Escherichia coli"
/clone lib="Aspergillus terreus random genomic DNA clone
library"
/notes="Vector: pZEROTM-2; Site1: Sau3A; Site2: BamHI;
Sau3A genomic fragments ligated into BamHI digested
pZEROTM-2 "
FEATURES
source

```

PZL-CAN-2					
ORIGIN					
Alignment Scores:					
Pred. No.:	3.36e-08	Length:	2121		
Score:	192.00	Matches:	90		
Percent Similarity:	37.80%	Conservative:	37		
Best Local Similarity:	26.79%	Mismatches:	103		
Query Match:	8.50%	Indels:	107		
DB:	28	Gaps:	15		
 US-09-985-689A-4 (1-433) x BZ424995 (1-2121)					
Qy	21	GlyClnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40		
Dd	1281	GGCGAGGCACCTACCGCTTACGTGATAGACCACCGCATCAACAGTGAC	1234		
Qy	41	SerMethisGluAlaPheArGgLyIsIleThrAlaLeuTyzrAlaLeuGlyArGthrAsn	60		
Dd	1233	-----CACGAGAAATTGAGGGCGGTGCGAGTCGCGTAC	1198		
Qy	61	AsnAlaser-----ApproAsnGlyHisGlyThrHisValAlaGlySer	75		
Dd	1197	AACGCTCCCGTGGGCCAGCATGTGGACGGTGTGGCCATGGCCCCACGTTTCTGGTACA	1138		
Qy	76	ValLeuGlyAsnAlaLeuAsnLysGlyMetAlaprogLnAlaAsnLeuValPheGlnser	95		
Dd	1137	ATTGCTGGTAGACATAC-----GGCGTGGCAAAGAGGCCAACCTGCTG	1090		
Qy	96	Ile-MetAspSerserGlyGlyLeuGlyLeuProSerAsnLeuAsn	111		
Dd	1089	GTCAAGCTCTCGTCGGGAATCAGCAGCACATCCATCATCTCTGGACGGCTTCAACTGG	1030		
Qy	112	-----ThreupheserGln-----AlaTr	118		
Dd	1029	GGCGCCNAATGACATGTGTTCCAGAAGCGCACCTGGAAAGGCGGATCAACATGAGCTTG	970		
Qy	118	pAsnAlaGlyAlaArgileHisThrAsenSerTrp-----GlyAlaPr	132		
Dd	969	GGTATGTTATCTCCCATCTATGTAATGA-TATCATTGGAGTCACACAGCTGACGCAAG	911		
Qy	132	oValasnGlyAlaTyrrThrAlaAsnSerArgGlnValAspGluTyrrValArgAsnAsnAs	152		
Dd	910	CATAGGCGGTGGAPACTCCAAGGCTTCAATGATGCCGTCCGAGAACCGCTTTCACAGAGGG	851		
Qy	152	pMerThrValLeupheAlaAlaGlyAsnGluGly-----ProAsnSe	166		
Dd	850	AGTCCTGTCCATCTGTCGCCGCGCAATGAGATGTGCGTGCCACGTTCCACCCCTGTGAA	791		
Qy	166	rGlyThrIleser-----AlaprogLy-T	174		

/note="Vector: Unizap; Metarhizium anisopliae sf. acridum was grown on insect cuticle and chitin for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."

ORIGIN

Alignment Scores:

Pred. No.:	5,15e-09	Length:	2141
Score:	200.00	Matches:	80
Percent Similarity:	43.75%	Conservative:	46
Best Local Similarity:	27.78%	Mismatches:	108
Query Match:	8.85%	Indels:	54
DB:	13	Gaps:	13

US-09-985-689A-4 (1-433) x BQ142519 (1-2141)

Qy	21	GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	1059	GGTGGAGGACTGCTGATATATATGACATGGTGTGAGGTCTCC	1012
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
Db	1011	-----CACCCGAGTTGGCGTGGCCCACTTGGCTCAGGAGTTCATCAACGGTCAA	958
Qy	61	AsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla	80
Db	957	AAC---CGTGATGGCCAGCCGACCTGAGTCACTCGCTGGTACTATT	907
Qy	81	LeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp	98
Db	906	CGAAGCTACGGTGTGCCAAAATGCCAAGCTTTTGTGTCAAGGTCTTGTATGACCAG	847
Qy	99	SerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp	118
Db	846	GGCAGTGGTCTCTACTCCGGTATCATCAGTGGCATGGACTTTGTT	796
Qy	119	AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsn	136
Db	795	GACTCCAGAGTCGTAACATGCCCAATGCCCACTTCTTCATGATGCTGGAGGTGGC	736
Qy	137	TyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAspMetThrValLeu	156
Db	735	TACTCGCGTCCGTCAACAGAGGTGCGGTGCTTTGGTCAGTCTGGTGTCTCTCTGCC	676
Qy	157	PheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLys	176
Db	675	GTCCCGCTGGACACGATACCGGATGCCCAACACCTCT	619
Qy	177	AsnAlaIleThrValGlyAlaThr	194
Db	618	ACTGCTGCACTTTGTCCTGCACTGCGTCAGATGACAGCGCATCTACCTTTTCCAACTAC	559
Qy	195	AlaAspAsnProAsnHisIleAlaGlnPheSerArgGlyAlaThrArgAspGlyArg	214
Db	558	-----GGCAQA	553
Qy	215	IleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla	234
Db	552	GTT---GTCGATATCTCGTCTCGTACCGGCATCTCTCCACC	511
Qy	235	ProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetClyGlyThrSer	254
Db	510	-----TGG-----ATCAATGGCCGCCCAACACCATCTCTGGCACCTCC	472
Qy	255	MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsn	274
Db	471	ATGGTACTCCCATATTGCTGT	424
Qy	275	ArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAsp	294
Db	423	AGCGGCAAGACTAGCCCTGCTCTTGTCCAGAGATCCAGGACACTTCTTACCAAGAAC	364
Qy	295	ValGlyLeuGlyTyrProAsnGly	302

Db	363	GTGATCCGCAATGTGCCCGTGGC	340
RESULT 8			
CA320325			
LOCUS			
DEFINITION	771 bp mRNA linear EST 09-JUL-2003		
	UI-M-FW0-cby-d-23-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone		
	IMAGE:6816072 5', mRNA sequence.		
ACCESSION	CA320325		
VERSION	CA320325.1	GI:24538449	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 771)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP)		
Seq primer:	pyX-5.		
Location/Qualifiers	1..771		
FEATURES			
source	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:6816072"		
	/tissue_type="whole brain"		
	/dev_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"		
	/lab_host="DH10B (T1 phage resistant)"		
	/clone_lib="NIH_BMAP_FW0"		
	/notes="Organ: Brain; Vector: pyX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System, supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."		
ORIGIN			
Alignment Scores:			
Pred. No.:	5.26e-09	Length:	771
Score:	192.50	Matches:	77
Percent Similarity:	39.86%	Conservative:	37
Best Local Similarity:	26.92%	Mismatches:	115
Query Match:	8.52%	Indels:	57
DB:	14	Gaps:	10
US-09-985-689A-4 (1-433) x CA320325 (1-771)			
Qy	8	ValLysAlaAspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlyGlnValAla	27
Db	40	CTGAGGAGATGTGCTGTGG---CAGATGGATACACAGGTGCTAATGTACAGAGTTGCT	96


```

DEFINITION      Fg02_10g08_R_Fg02_AAPC_EOORC_Fusarium_graminearum_mycelium
Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence.
ACCESSION       BI750157
VERSION         BI750157.1  GI:15771959
KEYWORDS        EST.
SOURCE          Gibberella zeae
ORGANISM        Gibberella zeae
REFERENCE       Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS        Hypocreales; Hypocreales; Nectriaceae; Gibberella.
1 (bases 1 to 718)
Harris, L.J., Glassco, T., Rocheleau, H., Allard, S., Chapados, J.,
Couroux, P., de Moors, A., Hattori, J.I., Ouellet, F., Robert, L.S.,
Singh, J.A., Spott, D. and Tinker, N.A.
Expressed Sequence Tags from Fusarium graminearum mycelium
Unpublished (2001)
Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris.lj@em.agr.ca.
FEATURES        Location/Qualifiers
source          1..718
                /organism="Gibberella zeae"
                /mol_type="mRNA"
                /strain="DAOM 180378"
                /db_xref="taxon:5518"
                /clone="Fg02_10g08"
                /tissue_type="Mycelial tissue"
                /dev_stage="Assexual"
                /lab_host="E. coli (Sure cells)"
                /clone_lib="Fg02_AAPC_EOORC_Fusarium_graminearum_mycelium"
                /note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
                Site 2: XhoI; Mycelial tissue was collected from V8 agar
                plates after a growth period of 6-7 days at 25 C with 14
                hrs (R/UV) day light exposure. Mycelia was ground in
                liquid nitrogen prior it's storage at -80 C until RNA
                extraction. Directional cloning with 5' end of cDNA cloned
                into EcoRI site of pBluescript and 3' end of cDNA cloned
                into XhoI site of pBluescript (Stratagene, La Jolla, CA)."
ORIGIN
Alignment Scores: 7.9e-10 Length: 718
Pred. No.: 200.00 Matches: 82
Score: 41.8% Conservative: 39
Percent Similarity: 28.3% Mismatches: 110
Best Local Similarity: 8.85% Indels: 58
Query Match: 12 Gaps: 12
DB:
US-09-985-689A-4 (1-433) x BI750157 (1-718)
QY 46 PheArgGlyLysIleThrAlaLeuTyxAlaLeuGlyxThrAsnAsnAlaSerAspPro 65
D 9 TTCGAGGTCGCTGCTCAGCTGTCTACACTCCCTTCAGCGGCCGCAAGCGT--GACACC 65
QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsnLysGlyMet 85
D 66 AACGTCACGGAACACGTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119
QY 86 AlaProGlnAlaAsnLeu-----ValPheGln---SerIleMetAspSerSer 100
D 120 GCCAAGAGGCCACCATCCAGCTGCAAGCTGCTCAGGCTCTCCAGGCTAGTTCATCCAGCACCTCC 179
QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla 120
D 180 ATCATCCTCGCTGGCTTCACTGGGCTGCACACGACATCATCTCCAG----- 227
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyxThrAlaAsn 140
D 228 ---GGCCGAACCAAGACACTCAGTCGTCATATGCTCTCTCGCGCGGTGCTACTCTGCTCTCC 284

```

```

QY 141 SerArgGlnValAspGluTyxValArgAsnAsnAspMetThrValLeuPheAlaAlaGly 160
D 285 TTCAACAACAGCTGTCGAGTCTKCVTCAGCTCCGGTATTATCTCTGTCATTGTCGGGT 344
QY 161 AsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
D 345 AACGATGGTGCACAGCTGCCAACACTCT--CTGCCCTCTCTCTCCAGCGCCAWCACT 401
QY 181 ValGlyAlaThrGluAsnTyxArgProSerPheGlySerIleAlaAspAsnProAsnHis 200
D 402 GTCGGTGC-----ATTGACAGCAACTGGGCC 428
QY 201 IleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThr 220
D 429 ATTTCCTCGTACTCCAACTACGCTACGCTCTC-----GATATCTTT 470
QY 221 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
D 471 GCYCTKGCACCAAGCTTCTCTCGGC-----TGG 500
QY 241 AlaAsnTyxAsnSerLysTyxAlaTyxMetGlyGlyThrSerMetAlaThrProIleVal 260
D 501 TACACCAACAGCAAGCTKCCACCAACCATCAGCGKACSTCCATGGCTACTCTCCACATT 560
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLys 280
D 561 GCCGAGCTTGTCTCKC-----TACGGTATTCTGTCAAGGTTGTTCT--GGT 605
QY 281 ProSerLeuLysAlaAlaLeuAlaGlyAlaThrAspValGlyLeuGlyTyxPro 300
D 606 GTCTCTTGTGTACCAACTGGCTGCTCTACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 647
QY 301 AsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyxVal 320
D 648 -----GGCAAGACACTGGCACTCCGCACTCCGCACTCCCACTCCCACTCCG 689
QY 321 AsnGluAlaThrAlaLeuAlaThrGly 329
D 690 CACAACGGCAGACAGCTTCAGTAAGCGT 716
RESULT 7
BQ142519/c
LOCUS          2141 bp      mRNA      linear      EST 24-APR-2002
Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium
anisopliae var. acridum cDNA, mRNA sequence.
ACCESSION      BQ142519.1  GI:20279578
VERSION        BQ142519
KEYWORDS        EST.
SOURCE          Metarhizium anisopliae var. acridum
ORGANISM        Metarhizium anisopliae var. acridum
                Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
                Hypocreales; Hypocreales; Clavicipitaceae; mitosporic
                Clavicipitaceae; Metarhizium.
REFERENCE       1 (bases 1 to 2141)
AUTHORS        Freimoser, F.M., Screen, S., Baga, S., Hu, G. and St. Leger, R.J.
TITLE          EST analysis of genes expressed by two different insect pathogenic
                fungi during optimized secretion of proteins
JOURNAL         Unpublished (2002)
COMMENT        Contact: Freimoser F. M.
                Department of Entomology
                University of Maryland
                4112 Plant Sciences Building, College Park, MD 20742, USA
                Tel: 301 405 16 13
                Fax: 301 314 92 90
                Email: ff34@umail.umd.edu.
                Location/Qualifiers
FEATURES        1..2141
source          /organism="Metarhizium anisopliae var. acridum"
                /mol_type="mRNA"
                /strain="ARSEF 324"
                /db_xref="taxon:92837"
                /clone_lib="Metarhizium anisopliae sf. acridum ARSEF 324"

```

```

Percent Similarity: 50.00%      Conservative: 40
Best Local Similarity: 31.13%    Mismatches: 69
Query Match: 9.40%              Indels: 37
DB: 12                          Gaps: 11

US-09-985-689A-4 (1-433) x BJ395336 (1-640)

QY      19  LeuTyrGlyGlnGlyGlnValValAlaValAlaLeuaspThrGlyLeuaspThrGlyArg--- 37
Db      22  TTAAGAGGTAAAGGTCAGATATTAGTATTGCTGATCTGGTTTAGATGGTAGCCATTGT 81
QY      38  -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48
Db      82  TTCCTTTTCAGATCCAAGTATCCAATACCATTTAATCAAGTGAATGAAATCATAGAAA 141
QY      49  LysIleThrAlaLeuTy-AlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHis 68
Db      142 GTTGTAACT-----TATATTACTTACCATGACATGAAGATTATTGTAATGGTCAT 192
QY      69  GlyThrHisValAlaGlySerValLeuGlyAsn-----AlaLeuAsn 82
Db      193 GGTACACAGTGTGTGGCTCTGCACAGCGGTACTCCAGAGGATTCCTCATGGGTATTTC 252
QY      83  -----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSer 100
Db      253 TCATTTAGTGGTCTTGCAACTCACGCAAGATTGCATTTATGATCTT-----TCATCT 306
QY      101 GlyGlyLeuGlyGlyLeu---ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsn 119
Db      307 GGAAGGTTCTGAACCAACACACCCCGAAGATTACAGTCAAATGTACAAACCATTTATGAT 366
QY      120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProValAsnGly 135
Db      367 GCAGGTGCAAGAGTACATGGTGATCTCTGGGGTCTGTATCTTTTGCAGAGGTATTATGGT 426
QY      136 AlaTyThrAlaAsnSerArgGlnValAspGluTyVal---ArgAsnAsnAspMetThr 154
Db      427 GGTATTTCGATGCTGGTGGTATTGATGCATTCCTCTATGATAGTACCCAGAATCTCT 486
QY      155 ValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThr 174
Db      487 ATACTAAGAGCTGCTGGTAAT---AACGAGCTATTTCATCTTTATTAGTCAAGCAACA 543
QY      175 AlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySer--- 193
Db      544 GCTAAAAATGCATTTACAGTTGGTGCTCGAGCAACAGCTCATGTAAATATTATGTGCAT 603
QY      194 -----IleAlaAspAsnProAsn 199
Db      604 GCATTGGAATATTATGATTTCTTCAGATTAATGCTAAT 639

```

TR319G10P	532 bp	DNA	linear	GSS 13-DEC-2000
LOCUS				
DEFINITION	T. Brucei sheared genomic DNA clone 319g10, forward sequence, genomic survey sequence.			
ACCESSION	AL492464			
VERSION	AL492464.1			
KEYWORDS	GSS.			
SOURCE	GI:11867408			
ORGANISM	Trypanosoma brucei			
	Trypanosoma brucei			
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;			
	Trypanosoma.			
REFERENCE	1. (bases 1 to 532)			
AUTHORS	Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Meiville,S.E., Rajadream,M.A. and Barrell,B.G.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhl@sanger.ac.uk			
COMMENT	Constructed at the Institute for Genomic Research (TIGR),			

Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (REU927/4 Gurat 10.1) was mechanically sheared to give a tight size distribution (1.5-4 kb). The $v+$ method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@uic.org
Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/projects/Tbrucei/>.

FEATURES
SOURCE

```
1.352
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319q10"
```

ORIGIN

Alignment Scores:	4.19e-10	Length:	532
Pred. No.:	200.50	Matches:	58
Score:	200.50	Conservative:	27
Percent Similarity:	46.20%	Mismatches:	72
Best Local Similarity:	31.55%	Indels:	2
Query Match:	8.87%	Gaps:	5
DB:	29		

US-09-985-689A-4 (1-433) X TA319G10P (1-532)

Qy	92	ValPheGlnSerIleMetAspSerSerGlyGlyLeuGly-----	104
Db	2	GTAATGCCAAGTATAATGGTGTCCACGGGGCGAAGATTCTT	61
Qy	105	-----GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrp	118
Db	62	CATCCACGTCAGAGCGTTGTCCTCCGCCACAGCGTTACTCAAA	121
Qy	119	AsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrp	138
Db	122	GCGCGTTCGAGCCCGTGTCTCTCAACTCGTGGGGTTTGTGCTCCCTCCGAGTATTC	181
Qy	139	AlaAsnSerArgGlnValAspGluTyrValArgAsn---AsnAspMetThrValLeuPhe	157
Db	182	GCTGTGGAAGGATATGGATGGTTGTCCGAGTAGTTATGACGATCGCTACTTATCTTC	241

RESULTS

REGULATORY
TA319G10P

INTEGRATIVE LOCUS

DEFINITION

1
2
3
4
5
6
7
8
9
10

ACCESSION

VERSION _____

KEYWORDS

SOURCE

ORGANISM

20190333

REFERENCE

AUTHORS

TITLE

JOURNAL

SECRET

COMMENT

COMMENT

```

QY 314 SerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThr 333
Db 1827 GCTTATCAGATCCTC-----AGCAGC 1847
QY 334 TyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSerLeu----- 348
Db 1848 TATAAACCGCAGGCAAGCCTG--AGTCCTAGCTACATCGACCTGACTGAGTGCCTAC 1904
QY 349 ValTrpThrAspAlaProGlySerThrThrAlaSerTyr-----ThrLeu 363
Db 1905 ATGTGG-----CCCTACTGCTCCAGCCTTACTACTATGAGGAATGCCAACATC 1955
QY 364 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAsp 383
Db 1956 GTTAAT-----GTCCATCTCTCAATGGCATGGCGCTCAGGAGAGAAATT 2000
QY 384 PheSerTyrProTyrAspAsnAsnTrp-----AspGlyArgAsnAsnVal 398
Db 2001 GTGGATAAGCCT-----GAGTGGCGACCTTATTACACAGAAATGGAGACAAT 2051
QY 399 GluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyrIleIleGlu 414
Db 2052 GAAGTGGCCTTCTCTACTCTCTCAGTGTGTGGCCCTGCTCAGGTTACCTTGCCATCTCC 2111
QY 415 Val 415
Db 2112 ATT 2114

RESULT 3
LOCUS BJ393752 594 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION BJ393752
VERSION BJ393752.1 GI:19304838
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 594)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..594
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores: 2.63e-11 Length: 594
Pred. No.: 213.00 Matches: 57
Score: 53.85% Conservative: 34
Percent Similarity: 33.73% Mismatches: 56
Best Local Similarity: 9.42% Indels: 22
Query Match: 12 Gaps: 9
DB:

US-09-985-689A-4 (1-433) x BJ393752 (1-594)

```

```

QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspPro---AsnGlyHisGlyThrHisVal 72
Db 33 TATATTACCACATCAACAAGCGACATAGTATAGTAAAGTGGATGTCACGGTACACATATT 92
QY 73 AlaGlySerValLeuGly-----AsnAlaLeuAsn-----LysGly 84
Db 93 TGTGTTCTCGCAGCAGGTACTCCAGAGANTTCTTCAGTTATATATTATTCATCTTAGTGGT 152
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyGlyLeuGly 104
Db 153 CTTCGAACGTGATGCAAGAGATTGTCATTCTTTGATTGGCAAGTGGTTCATCAAGTTTGACA 212
QY 105 GlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyValaArgIle 124
Db 213 CCT---CATCGGATTTGAAACAATTATATCAACATTATATGACGAGGTGCAAGAGTG 269
QY 125 HisThrAsnSerTrpGlyAla-----ProValAsnGlyAlaTyrThrAlaAsn 140
Db 270 CATTTGATTCTTGGGGTTCTGTATCAGTAGAGGGGTATACAGGTAGTTATTTCATCAGAC 329
QY 141 SerArgGlnValAspGluTyrVal---ArgAsnAsnAspMetThrValLeuPheAlaAla 159
Db 330 ACTGCTTCAATGTGATGATTCCTTTTCACTCATCCAGATTTTCATCTTTAGAGCTGCT 389
QY 160 GlyAsnGluGlyProAsnSerGlyThrIleSer-----AlaProGlyThrAlaLysAsn 177
Db 390 GGTAAAC-----AACGACGAATACCTATCACTCACTCACTCAATCCAGTCAAAGAAT 440
QY 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsn 197
Db 441 GTTATTACCGTTGGTGCTCATCAACAACATTCATCAAAATTTAT-----TTAACTGATGGT 494
QY 198 ProAsnHisIleAlaGlnPheSerSer 206
Db 495 CCATAATTATATAAATTATCAATCATCT 521

RESULT 4
LOCUS BJ395336 640 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds38b16 5', mRNA sequence.
ACCESSION BJ395336
VERSION BJ395336.1 GI:19306422
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 640)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..640
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores: 3.37e-11 Length: 640
Pred. No.: 212.50 Matches: 66
Score:

```

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muranatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9226, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
URL: http://Location/Qualifiers

FEATURES

source

```
1. .4198
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FANTOM_DB:4732484M11"
/db_xref="MGI:2391080"
/db_xref="taxon:10090"
/clone="4732484M11"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
417. .3575
/notes="unnamed protein product; membrane-bound transcription factor protease, site 1 (MGJ|MGJ:1927235, GB|NM_019709, evidence: BLASTN, 99%, match=3782) putative"
/codon_start=1
/protein_id="BAC26263.1"
/db_xref="GI:26325018"
/translation="MKLVSTWLLVLLVLCGRHLGDRIGTRALSKAPCPSCSHLTKE VEFSTVVEYEVIVAFNGFYTAARNSEFISSALKSEVENWRIIPRNPSSDYDFE VIQIKKQAGLITLDEHFNIXRVTPQKVFSLKFAENPIVPCNTRWSQKQWSSR PLKRALSLGSGFHWATGRSSRRLLRAIPROVATLOADVLMQGYTGANVRVAVFD TGLSEKHPKXVKERTNTNERTLDDGLGHGTFVAGVTASRECGGEAPDAELHIFR VFTNQSVTSWFLDAPNYALLKMDVNLNLSGDPDMDFVDKXWELTANNVMS AIGNDPLGLTNLPADQMDVIGVGIDFEDNIARFSSRGMTWELPGYGRKFDIV TYAGVRSVGKGCALSGTIVASPVVAGAVTLVSTVQKRELNPASVXOALIASA RLPGVNFQHGKLDLLRAYQILSSYKPSLSPFSDIDTECPYMPYCSQPIYQG GMPTIVNVITLNGMGVTRIVDKPEWRPYLPQNGNIEVAFYSVSLWPSGYLAISI SVTKAASWEGIAQGHIMITVASPAETELHSGAHTSTVKLPKVKIIPTPERSKRLV WQGYHNLRYPGVPEEDNLRMKNDPLDWDGHDVHTNFRDMYOHLSMSGVEVELCAPF TCFDQYGTLLLVDSSEYFPREIAKLBDVDNGLSLVIFSDWNTYMSRKVYFDE NTRQWMPDITGANIPALNELLSSVNMWFSDEGLYEFGEFVLANHNTYASGCSIAKFE DGVTITQTFDQGLEVLKQETAVENVPLIGLYIPSEGGGRIVLYGNSNCLDDSHRQ KQCFWLLDALLQYTSYGVTPSPSLSGNRQRPSPGAGLAPPERMEGNHLHRYSKYLEA HLGDPKPRPLPACFLHSWAKPOLNLTAPSNLWIKQLISIDLVKVLNPNFRSNRPQV RPLSPGSGAWNDIPGGIMPGRYNOEVGQTIPIVFAFLGMAVALAFVFWQISAKSRPKR RRPRAKRPOLAQAHPARTPSV"
```

CDS

polyA_signal
polyA_site
ORIGIN

Alignment Scores:

Pred. No.:	5.61e-11	Length:	4198
Score:	224.00	Matches:	113
Percent Similarity:	38.32%	Conservative:	56
Best Local Similarity:	25.62%	Mismatches:	168
Query Match:	9.91%	Indels:	104
DB:	11	Gaps:	21

US-09-985-689A-4 (1-433) x AK029048 (1-4198)

Qy	8	VallyysalaasppvalalalaglnaasentyrGlyLeuTyrglyGlnValValala	27
Db	1005	CTGCAGGCAGATGCTGTGG---CAGATGGGATACACAGTGTCTATGTGCAGAGTTGCT	1061
Qy	28	ValalaasppThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Db	1062	GTTTTTCATACTGGCTC-----AGTCAGAAGCATCCGCATTTTAAG	1103
Qy	48	GlylysileThrAlaLeuTyralaLeuGlyArgThrAsn-----AsnAla	62
Db	1104	AAT-----GTCAAGGAGAGAACCAACTGGACCAATAGACGGACCGCTG	1145
Qy	63	SerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaLeuAsn	82
Db	1146	GATCATGGGTAGGCCATGGCACATTGCTTCAGGTGTGATGCCACATCAGGAGTGC	1205
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerSerGly	101
Db	1206	CAAGGATTGCTCCAGATGCAGAGTGCACATCTTCAGGGTCTTTACCAACAATCAG--	1262
Qy	102	GlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTatPAsnAlaGly	121
Db	1263	-----GTGCTTACATCTGTTTCTGGATGCTTCACTATGCC	1304
Qy	122	AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsnSer	141
Db	1305	ATCCTAAAGAAGATGGACGCTTCTCAACCTTAGCATCGTGGCGCCGACTTCATGGATCAT	1364
Qy	142	ArgGlnValAspGluTyr-----ValArgAsnAsnAspMetThrValLeuPheAlaAla	159
Db	1365	CCGTTTGTGACAAAGGTGGGAATTAACAGCTAAACAATATTAATATGGTTCTGCTATT	1424
Qy	160	GlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaIleAsnAlaIle	179
Db	1425	GGCAATGATGCACCTCTCTATGGCACTCTGAATAACCTGCTGATCAGATGGATGTGATT	1484
Qy	180	ThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsn	199
Db	1485	GGAGTGGGTGGC-----ATTGACTTTGAAGAT	1511
Qy	200	HisIleAlaGlnPheSerSerArgGlyValaThrArg-----AspGly	213
Db	1512	AACATCGCTCGCTTTCTTCCAGGGGAATGACTACCTGGGAATTACCAGGAGGCTATGGT	1571
Qy	214	ArgIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu	233
Db	1572	CGTGTGAAGCTGACATGTGCACCTAGTGGTGGTGGAGTGGCGGGT-----	1616
Qy	234	AlaProAspSerSerPheTyrAlaAsnTyrAsnSerLysTyrAlaTyMetGlyGlyThr	253
Db	1617	-----TCGGTGTGAAGAGGGGGCTGCCGTGCACCTCTCAGGAGACC	1655
Qy	254	SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys	273
Db	1656	AGTGTGCTTCCCAGTGGTGGTGGCGGCTCACCTTGTGTAGTAAGCACAGTACAGAAG	1715
Qy	274	AsnArgGlyIleThrProLysProSerLeuIleLysAlaAlaLeuIleAlaGlyAlaThr	293
Db	1716	CGGAGCTGGTG-----AATCTCGCAGTGTGAGCAAGCTTTGATAGCGTACGCCGG	1769
Qy	294	AspValGlyLeuGlyTyrProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLys	313
Db	1770	AGACTTCTCT---GGGGTCAACATCTTCAGCAAGGTTCATGCAAGTTGGATCTCTCTCGCA	1826

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
 TITLE EST analysis of genes expressed by the zygomycete pathogen
 Conidiobolus coronatus during optimized secretion of proteins
 JOURNAL Unpublished (2002)
 COMMENT Contact: Freimoser F.M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@umd.edu.

FEATURES
 source
 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus cornatus ARSEF 512"
 /notes="Vector: UniZap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector UniZap."

ORIGIN
 Alignment Scores:
 Pred. No.: 5,6e-13 Length: 1605
 Score: 236.50 Matches: 88
 Percent Similarity: 43.75% Conservative: 38
 Best Local Similarity: 30.56% Mismatches: 87
 Query Match: 10.46% Indels: 75
 DB: 13 Gaps: 14

US-09-985-689A-4 (1-433) x BQ622771 (1-1605)

Qy 6 GlyIleValIysAlaAspValAlaGlnAsnAsn-----TyrGlyLeu----- 19
 Db 718 GGTGTTCTTAAGCTATGCGGTCCAGCAAGCAATGCTCTGGGTCTGCTGTTGGT 777
 Qy 20 -----TyrGlyGlnGly 23
 Db 778 CAACGTGCTAAGCTCGGATCTGCTCTTACTTACCAACCATGCTGCTGCTCAAGT 837
 Qy 24 GlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHis 43
 Db 838 GTCACTGTTTTCGTTTATGATCTGCTGCTCAAGCAAGCAATGCTGCTGCTGCTGCT 885
 Qy 44 GluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAla--- 62
 Db 886 -----TTGCTGCTGCTGCTGCTGCT-----TGGGTACTAACAACCTGCTGCT 924
 Qy 63 -----SerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 925 GGTAGCAACACTGATGCTCAGGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984
 Qy 80 AlaLeuAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet----- 97
 Db 985 ACCTAT-----GGTGTGCAAGAGGCTAACATGCTGCGTTAAGGCTTAGGTGAT 1038
 Qy 98 AspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
 Db 1039 GATCGTCCGGATCATCTCTGGAATATCTCCGGTATTCCTGCTGCTGCTGCTGCTGCTGCT 1095
 Qy 118 TrpAsnAlaGlyAlaAspGileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyr 137
 Db 1096 ---TCTGCTGCCAAGAAAGTTATCTCTATGATTAGGAGTGTGTAAGACGATGCTCTT 1152
 Qy 138 ThrAlaAsnSerArgGlnValAspCluTyrValArgAsnAsnAspMetThrValLeuPhe 157
 Db 1153 -----AACACTGCTGTTTAAACAAGCTCTCAGCAAGGAGGTGCTGCTGCTGCTGCT 1200
 Qy 158 AlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsn 177
 Db 1201 GCTGCTGCTAAGCAACAGAGATGCTGTGGATCTCT---CCCGCTTCTGCTCTCTCC 1257

Qy 178 AlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsn 197
 Db 1258 GCCATTACCGTGTGCTGCTGCT-----GATGTC 1284
 Qy 198 ProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysPro 217
 Db 1285 AATGATAAAAGGCTTCACTTCTTAACCTCGGTAGCTGT-----GTC 1326
 Qy 218 AspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 237
 Db 1327 GATATCTTAGCTCTGCTGCTGCTCAACATCTCTCCACC----- 1362
 Qy 238 SerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThr 257
 Db 1363 -----TGGAAGGATCTAACACTGCCACCAACACATCTCTGCTACTCTGCTGCTGCT 1416
 Qy 258 ProfileValAlaGlyAsnValAla 285
 Db 1417 CCTCACATTGCTGCTGCTGCTGCTGCT 1440

RESULT 2
 AK029048
 LOCUS Mus musculus 4198 bp TRNA linear HTC 18-SEP-2003
 DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
 library, clone:473484M11 product:membrane-bound transcription
 factor protease, site 1, full insert sequence.
 ACCESSION AK029048
 VERSION AK029048.1 GI:26325017
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muranatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

3
 Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, J.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muranatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861

4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 5 (bases 1 to 4198)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-4
Perfect score: 2260
Sequence: 1 NDVARGIKADVQAQNNYGLY.....EVQAYNVSPGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat_31032004_161807_4168/app_query.fasta_1.3498
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=prt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CGN_1_1_7084 @runat_31032004_161807_4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcm:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	236.5	10.5	1605	13	BQ622771 CC_Contig
2	224	9.9	4198	11	AK029048 Mus muscu
3	213	9.4	594	12	BJ33752 BJ393752
4	212.5	9.4	640	12	BJ395336 BJ395336
5	200.5	8.9	532	29	TA319G10P
6	200	8.8	718	12	BI750157
7	200	8.8	2141	13	BQ142519 Contig56 M
8	192.5	8.5	771	14	CA320325 UI-M-FW0
9	192	8.5	2121	28	BZ424995 100023066
10	191.5	8.5	601	12	BJ387574 BJ387574
11	190	8.4	545	13	BU575479 TGESTyB8
12	189	8.4	614	9	AJ273402 AJ273402
13	188	8.3	574	29	TA315H10P
14	185.5	8.2	716	28	BZ833395 HU2_0177
15	181.5	8.0	665	13	BQ770462 UI-M-F10
16	181	8.0	650	9	AJ274038 AJ274038
17	180	8.0	530	29	CNS010EO
18	179	7.9	1029	29	CNS071DW
19	178.5	7.9	675	14	CF727824 UI-M-HB0
20	178	7.9	508	28	AO652212 Sheared D
21	178	7.9	633	12	BJ369190
22	177.5	7.9	576	14	CD235943 ScrP691
23	177.5	7.9	641	12	BU393925 BJ393925
24	177.5	7.9	1002	29	CNS06DGB
25	176.5	7.8	594	14	CF846026 psHB036xA
26	176	7.8	580	9	AJ273745 AJ273745
27	176	7.8	583	9	AJ273947 AJ273947
28	176	7.8	593	9	AJ273918 AJ273918
29	176	7.8	601	9	AJ273050 AJ273050
30	176	7.8	601	9	AJ273921 AJ273921
31	173	7.7	610	9	AJ274218 AJ274218
32	170	7.5	609	9	AJ272712 AJ272712
33	170	7.5	681	14	CB690041 CBST-54-B
34	169	7.5	564	28	AO651427 Sheared D
35	168.5	7.5	615	9	AJ274059 AJ274059
36	168	7.4	794	14	CF737198 UI-M-HD0
37	167.5	7.4	937	12	BG246418 602360428
38	167	7.4	781	13	BU612128 UI-M-EW0
39	166.5	7.4	593	9	AJ273903 AJ273903
40	166.5	7.4	596	9	AJ273379 AJ273379
41	166.5	7.4	619	9	AJ273097 AJ273097
42	166	7.3	887	13	BQ879057 AGENCOURT
43	165	7.3	604	9	AJ273185 AJ273185
44	165	7.3	1572	11	AY107161 Zea mays
45	164.5	7.3	604	9	AJ274126 AJ274126

ALIGNMENTS

RESULT 1
BQ622771
LOCUS CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
DEFINITION CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
ACCESSION BQ622771
VERSION BQ622771.1
KEYWORDS EST.
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

Job time : 7336.25 secs

```
Db      3007 GCCTAC-----GGCTGCCCAAGAGGCCAAGATCGTAGGGCTCGCGTCCGGTGTGTAACAAC 2954
Qy      100 Ser-----GlyGlyLeuGlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db      2953 TCCGGCCAGGACACACCCCGAGTCTGTCGGCGGATCGACTGGTCTCGCCGGAACGCC 2894
Qy      118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db      2893 GT-CAAGCCGGCGCTCGCCACAT-----GTCCCTCGGCGGCGGCGCGCA 2850
Qy      137 rThrAlaAsnSerArgGlnValAspGluTyValArgAsn-----AsnAspMetTh 154
Db      2849 CACGGC-----CTGCACACGGCGGTACGCAACGCCATGGCTCCGGCGTCAC 2802
Qy      154 rValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db      2801 CTTCCGCGTGGCGCGGCAACAGTCCGACCAACGCTCCACGAGGTCA---CCCGCAG 2745
Qy      174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyArgProSerPheGlySerI 194
Db      2744 CGTCACGGAGGCCATACGGTCCGGCGGACGACCTCGGACGCCAAGCCGCGC--- 2690
Qy      194 eAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db      2689 -----TACTCCAATACGGCTCCGCTCTC----- 2666
Qy      214 gIleLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db      2665 -----GACCTCTTCGCCCGCGGTTCGTCCATCAGCTCGGCC----- 2630
Qy      234 aProAspSerPheTrpAlaAsnTyAsnSerLysTyAlaTyMetGlyGlyThrSe 254
Db      2629 -----TGGAACTCAAGCGACTCGGCGGACCAACACCATCTCCGGTACGTC 2586
Qy      254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db      2585 GATGGCACCACCGCACCTGGCGGCGCGCGCGCTC-----CACCTGCCGCCAA 2535
Qy      274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaG 291
Db      2534 CCCCTCGGCACCCCGTCCAGGTCCGCACGGCGCTGACGTCCGCGCCACACCGCGGT 2475
Qy      291 yAlaThrAspValGlyLeuGlyTyProAsnGlyAspGlnGlyTrpGlyArgValThrLe 311
Db      2474 CGTCACCAACCCCGCGCGGCTCGCCCAAC----- 2444
Qy      311 uAsnLysSerLeuAsnValAlaTyValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
Db      2443 -----CGGTCCTGTACGTCCGCGCGGCGCACACCCCTCCGGCGCGCG 2397
Qy      331 s-----AlaThrTySerPheGlnAlaGlnAlaGlyLysProLeuLy 345
Db      2396 CTTCCGAGAACACCGGTGACTACACGATCAGCGACAACTCCACGGTCGAGTCCCGGTGAC 2337
Qy      345 sIleSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db      2336 GGTCTCCGGCTCTCCGGCAACGCGGCTCGGCCCTCGCGCTAGAGGTCCACATCGTCCA 2277
Qy      361 rThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyValG 381
Db      2276 CACGTACATCGCGACCTCCAGGTCCAGTCTATCCGCCCGGACGCGCACGGGTACAGCT 2217
Qy      381 yAsnAspPheSerTyProTyAspAsnAsnTrpAsnGlyArgAsnAsnValGluAsnVa 401
Db      2216 CAAG-----TCGTACGCGCACCGGCGAGTTCGGACAACATCAACACCGTACTCGGT 2163
Qy      401 lPheIleAsnAlaProGln-----SerGlyThrTyIleIleGluVal 415
Db      2162 G-----AACGCTCTCGGAGGCGGCCAACGCGACGTGGAACTCGGGGTG 2117
```

Search completed: April 4, 2004, 23:08:56

[illegible]

Db 2693 CTGAAGCCTACAGCAGCTTCTCGGC 2719
RESULT 13
US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113
Alignment Scores:
Pred. No.: 6,98e-23 Length: 1560
Score: 298.00 Matches: 126
Percent Similarity: 42.11% Conservative: 50
Best Local Similarity: 30.14% Mismatches: 153
Query Match: 13.19% Indels: 90
DB: 15 Gaps: 19
US-09-985-689A-4 (1-433) x US-10-084-846A-113 (1-1560)
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTGACGGGTACGTATCGACACCGCGTC-----CGC 492
Qy 41 SerMetHisGluAlaPheArgGlyHisLeuThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCCACACAGCGACTTCGGGGCGGGCTCTTACCGGTACGACGCGCATCGACACGAC 552
Qy 60 AsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCAGGACGGCCACGGCCACGGCACGCGTGGCGCGCACGGTCCGGCGCAAC 612
Qy 80 AlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 613 GCCTAC-----GGCGTCGCAAGAGGCAAGTCTAGCGTCCCGGTGCTGACAAAC 666
Qy 100 Ser-----GlyGlyLeuGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAla 117
Db 667 TCCGGCCAGGACACCGCCAGGTCTCGCCGGCATCGACTGGGTGGCGCGCAACGCC 726
Qy 118 TrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro-ValAsnGlyAlaTy 137
Db 727 GT-CAAGCGCGCGCTCGCAACAT-----GTCCCTCGCGCGCGCGCGCGA 770
Qy 137 rThrAlaAsnSerArgGlnValAspGluTyrValArgAsn-----AsnAspMetTh 154
Db 771 CACGGCC-----CTCGACCGCGCGTACGCAACGCCATGCGCTCCCGCGCTCAC 818
Qy 154 rValLeuPheAlaGlyAsnGlyProAsnSerGlyThrIleSerAlaProGlyTh 174
Db 819 CTTCGCGCGCGCGCGCAACGAGTCGACCAACGCGCTCCACGAGGTCA---CCGCGACG 875

Qy 174 rAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIle 194
Db 876 CGTACCCAGGCGCATCACGCTCGGCGGACGACGAGCTCGGACGCCAAGCCGCGC-----930
Qy 194 eAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyAr 214
Db 931 -----TACTCCAACTACGGCTCCGCTCTC-----954
Qy 214 gileLysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAl 234
Db 955 -----GACCTCTTCGCCCGCGGTTCGTCATCATCACCTCGGCC-----990
Qy 234 aProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThrSe 254
Db 991 -----TGGAACTCAAGCGACTCGGCGACCAACACCACTCTCCGGTACGTC 1034
Qy 254 rMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAs 274
Db 1035 CATGGCGACCCCGACGCTGGGGCGCGCGCGCTC-----CACCTCGCGCGCAA 1085
Qy 274 nArgGlyIleThrProLys-----ProSerLeuIleLysAlaAlaLeuIleAlaGl 291
Db 1086 CCCCTCGGCCACCGCGTCCCGAGGTGGCGCGCTGACGTCCGCCGCCACACCGCGCT 1145
Qy 291 yAlaThrAspValGlyLeuGlyTyrProAsnGlyAspGlnGlyTrpGlyArgValThrLe 311
Db 1146 CGTACACCAACCCCGACGGGCTCGCCAAC-----1176
Qy 311 uAsnLysSerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLy 331
Db 1177 -----CGGCTCTCTGTACTCGGCGCGCGCGCGACACCACTCTCCGGCGCGCG 1223
Qy 331 s-----AlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLy 345
Db 1224 CTTGAGAACACCGGTGACTACAGATCAGCGACAACTCCACGGTCGAGTCCCGGTGAC 1283
Qy 345 sileSerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----Ty 361
Db 1284 GGTCTCCGGCGTCTCCGGCAACGGCGCTCGCGCTCGCGTAGAGGTCCACATCGTCCA 1343
Qy 361 rThrLeuValAsnAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGl 381
Db 1344 CACGTACATCGCGACCTCCAGGTCCAGTATCGCCCCCGACGCGCGGTACAGCT 1403
Qy 381 yAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVa 401
Db 1404 CMAA-----TCGTACGGCACCGCGCGCGGTTCGGACACATCAACACCATCTCGT 1457
Qy 401 lPheIleAsnAlaProGln-----SerGlyThrTyrIleIleGluVal 415
Db 1458 G-----AACGCTCTCTCGGAGCGCGCACACGCGACGTGGAACTGCGGGTG 1503

RESULT 14
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816

QY 377 -----GlnIleTyrValGlyAsn-----AspPheSer 385
Db 2221 CCGTGAATTGTCAGTGGCACATTAAAGTACGTAGTGGGACACCGAGTACAGAACTTTTGG 2280
QY 386 TyrProTyrAspAsnAsnTrp-----AspGly-----ArgAsn 396
Db 2281 ATCTATGCAACTGAGCCATGGATTAGCCCTTTTGTGAGTGAAGTGTAAATCTAGAGAAC 2340
QY 397 AsnValGluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyr 411
Db 2341 AATACCGAGTTTGTCTTATAGGTGAAATATGATAGAGGGTCTTGGACCGAGTCTCTAT 2400
QY 412 -----IleIleGluValGlnAlaTyrAsn 419
Db 2401 GTTGAAGGATAATCATTTGATGATCCACACGCCAGTTATTGAAGACGAGATCTTGAAC 2460
QY 420 ValProSerGlyProGlnArgPheSer 428
Db 2461 ACAATTGTTATTCCCGAAGAATTCACT 2487

RESULT 12

US-09-927-827-29
; Sequence 29, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseyer, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 29
; LENGTH: 3743
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2743)
US-09-927-827-29

Alignment Scores:

Score: 1,32e-22 Length: 3743
Percent Similarity: 30.50 Matches: 138
Best Local Similarity: 40.51% Conservative: 52
Query Match: 13.30% Mismatches: 150
DB: 10 Indels: 129
Gaps: 26

US-09-985-689a-4 (1-433) x US-09-927-827-29 (1-3743)

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeu-----AspThrGly 36
Db 1502 GGCAGCGGCACGCTGCTGGCGGTGATTGATACCGCATCACCAGTCATCCGACCTCAAC 1561
QY 37 ArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeu 56
Db 1562 GCCAACATCTCCGGGGTACGACTTCATCAGCGATGCCAGCCACCGCATGCCATGCCAAC 1621
QY 57 GlyArgThrAsnAsnAlaSerAsp-----ProAsnGlyHisGlyThrHisValAlaGlySer 64
Db 1622 GGCCTGTGACAGAACACGCCCGCAGGAGCGACTGGTGCAGCGCCAAACGAATGCCGCGCC 1681
QY 65 -----ProAsnGlyHisGlyThrHisValAlaGlySer----- 75
Db 1682 GGCATTCCCGCCGCCAGCTCCAGCTGGCAGCGGACCCATGTGGCCGGCAGCGGTTCGGGCA 1741
QY 76 -----ValLeuGlyAsnAlaLeuAsnLysGlyMetAlaProGln 88
Db 1742 GTGACCAACACACACCACCGCGGTAGCCGCGACCGCTACGCGCGCAAGTCTGTCGGGTG 1801

QY 89 AlaAsnLeu-----ValPheGlnSerIleMetAspSer 99
Db 1802 CGGTGCTCGGCAAGTCCGTCGGTGCCTCGGATATCCCGACGCGCATTCGTGGGCC 1861
QY 100 SerGlyGly-----LeuGlyGlyLeuProSerAsnLeuAsn-----ThrLeuPheSerGln 116
Db 1862 TCCGGCGCACCGTCCAGCGCATCCCGGCCAATGCTAACCCGCGGAGGTGATCAACATG 1921
QY 117 AlaTrpAsnAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 136
Db 1922 TCGTTCGGCGCGCGGTAGTCTGCTGACCAACCCATGCGGCAACGCC-----ATCAACGGTGGC 1978
QY 137 TyrThrAlaAsnSerArgGlnValAspGluTyrValArgAsnAsnAspMetThrValLeu 156
Db 1979 GTGTCGCGCGCGCAC-----ACGGTGGTG 2002
QY 157 PheAlaAlaGlyAsnGluGlyProAsn-----SerGlyThrIleSerAlaProGlyThrAla 175
Db 2003 GTCGCGCGCGCACGATGCGTCCATGTCGCGT-----TCGCTGCCGCGCAACTGC 2056
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsnTyrArgProSerPheGlySerIleAla 195
Db 2057 GCGAACGTGATTCGGTGGCGCGCAC-----ACCTCGCGCGCGCG----- 2098
QY 196 AspAsnProAsnHisIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIle 215
Db 2099 -----AAGGCGCAGCTATTCCACTTCGCGCAC----- 2125
QY 216 LysProAspValThrAlaProGlyThrPheIleLeuSerAlaArgSerSer-----Leu 233
Db 2126 GGTATCGATGTGTCGGCGCGCGGTCTGCTGATCTCTCCAGCTCAACACGCGGCACACC 2185
QY 234 AlaProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyrAlaTyrMetGlyGlyThr 253
Db 2186 ACGCCGCGTAGCGCAGCTATCTCTCAAC-----GGCAC 2224
QY 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 273
Db 2225 TCGATGGCGTGGCGCATGTGGCGCGGTGTCGCTGCGTGTGCAGTCG-----GTGCC 2278
QY 274 AsnArgGlyIleThrProLys-----ProSerIleuLys-----AlaAlaLeuIle 289
Db 2279 CCGACCGCGTGCAGCGCAGCGCGGTGGAACCTTGTGAAGAACACCGCGCTGCTTTA 2338
QY 290 AlaGlyAlaThrAspValGlyLeuGlyTyrPro----- 300
Db 2339 CCGGCGCGCTGCTCGGCGCGGTGCGGTGCGCGCATCTCAACCGCGGTGCGCGGTCACT 2398
QY 301 -----AsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnVal 317
Db 2399 CGCGCCATCAATGCGGAGCGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2458
QY 318 AlaTyrValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGln 337
Db 2459 ACTCCGCTGACCGCGCTGGCGCG-----GCGACTGGCGCGGAATGAATACCATCACC 2515
QY 338 AlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr 357
Db 2516 GTGCGCGCGCGCGCG-----ACCTGACGCTGACCGCGCGCGCGCGCGCGCGCGCGCG 2563
QY 358 ThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIle-----ThrAlaPro 374
Db 2564 -----GGCATGCGCGACCTGTATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2602
QY 375 AsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGly 394
Db 2603 ACCGACTCGGCTACAC-----TGGCGCCCATACCGCAG----- 2638
QY 395 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrIleIleGlu 414
Db 2639 ---GGCAATGCGGAGACTGCACCATCAGCGCACCG---TCCGGAAGTATTAGTCGCT 2692
QY 415 ValGlnAlaTyrAsnValProSerGly 423


```
Qy 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 1027 GAGTGGCGGCGCCGAGCAGCGCGGACGTCGTCACACCTGAGCTGGCGCGGCGACACC 1086
Qy 123 gileHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgGl 143
Db 1087 CCGGAGATCGACCCGCTGGAA-CGGAGAGTCAACAGCTC----- 1125
Qy 143 nValAspGluTyValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 1126 -----TCCGAGGAGAGGCGATCCTCTTCGCGATCGCGCGCGCAA 1166
Qy 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 1167 CGAAGCGAGTTCGGCGAGCAGACCATCGCTCCCGCGCAGCGCGCGCGCTCAC 1226
Qy 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 1227 CTTCCGCGGCC----- 1253
Qy 200 sIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 1254 GCTGGCGTCTCTTCAGCGCGCGCGCGCGCTGGAGCGCGCCATCAAGCCGCGCTCAC 1313
Qy 220 rAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPheTr 240
Db 1314 CGACCCGCGCTGCATCACCGCGCC-----GCGGCGCGCGCGCTCATCGA 1364
Qy 240 pAlaAsnTyAsnSerIlys-----TyrAlaTyMetGlyThrSerMetAlaTh 257
Db 1365 CCAGGAGTTCGGCGCAGAGCGCGCTACCTCACATCTCCGCTACGCTCGATCGCGAC 1424
Qy 257 rProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyI 277
Db 1425 CCGCATGTCGCGCGCGCGCGCGCTCAAGCGCGCTCAAGCAGCAGC----- 1467
Qy 277 eThrProLysProSerIleu-----IleLysAlaAlaLeuIleAlaGlyAlaThrAspVa 295
Db 1468 -----CCCAACTGCTCTCGCGCGCTCAAGCGCGCTCAAGCGCGCTCGCGAAG----- 1518
Qy 295 lGlyLeuGlyTyTrpProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLe 315
Db 1519 -GGCGGCAAGTACACCGCTTCAGCAGCGCGCTCGGCGCGCTATCGCGCTCGACCAAGCGAT 1577
Qy 315 uAsnValAlaTyValAsnGluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTy-Se 335
Db 1578 CAAGCAGTCGCTCATCGCAACCCGAACCTCGGTGAGCTTCGGCAT-CCAGCAGTGGCGCG 1636
Qy 335 rPheGlnAlaGlnAlaGlyLysPro 343
Db 1637 ACACGACGACAGCGCGCTCAOCCA 1661

RESULT 10
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication NO: US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HAITORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
```

```
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 2,21e-20 Length: 9025608
Score: 325.00 Matches: 113
Percent Similarity: 45.56% Conservative: 46
Best Local Similarity: 32.38% Mismatches: 125
Query Match: 14.38% Indels: 67
DB: Gaps: 12
US-09-985-689A-4 (1-433) x US-10-156-761-1 (1-9025608)
Qy 18 GlyLeuTyTrpGlyGlnGlyGlnValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 4132505 GGGTACGACGCGCAAGGCGCTCAAGATCGCGTCTTGACACCGGTGTGCAC----- 4132455
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57
Db 4132454 -----GCGACCCACCGGACCTCAAGACAGGTGGCGGAGTCCCAAGAACTTCTCC 4132404
Qy 58 ArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 4132403 GCCCGCGCGCGCGCGCGCGCTTCGTCAGCGCAGCAGCGCTCGCTCGCGCGCG 4132344
Qy 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 4132343 GGCACCGCGCGCGCGCGCGCTCAAGCGAGTACAGGTGTCCGCGCGCGCGAGATCTCTC 4132284
Qy 93 PheGlnSerIleXetAspSerSerGly-----GlyLeu 103
Db 4132283 AACGCCAAGTCTCTCGACGACACCGGCTCCGCGCGAGCTCCGCGCATCTTGGCGCGATG 4132224
Qy 104 -GlyGlyLeuProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaAr 123
Db 4132223 GAGTGGCGCGCGCGCGCGCGCTCGTCAACCTGAGCTGGCGCGCGCGACACC 4132164
Qy 123 gIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAlaAsnSerArgGl 143
Db 4132163 CCGGAGATCGACCCGCTGGAA-CGGAGAGTCAACAGCTC----- 4132125
Qy 143 nValAspGluTyValArgAsnAsnAspMetThrValLeuPhe-----AlaAlaGlyAs 161
Db 4132124 -----TCCGAGGAGAGGCGATCCTCTTCGCGATCGCGCGCGCAA 4132084
Qy 161 nGluGlyPro---AsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleTh 180
Db 4132083 CGAAGCGGAGTTCGGCGAGCAGACCATCGCTCCCGCGCGCGCGCGCGCGCTCAC 4132024
Qy 180 rValGlyAlaThrGluAsnTyArgProSerPheGlySerIleAlaAspAsnProAsnHi 200
Db 4132023 CGTCCGCGGCC-----GTGAACGACGCGCAA 4131997
Qy 200 sIleAlaGlnPheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValTh 220
Db 4131996 GCTGGCGTCTCTTCAGCGCGCGCGCGCGCTCGAGCGCGCGCATCAAGCCGCGCTCAC 4131937
Qy 220 rAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTr 240
Db 4131936 CGCACCGCGCGGTGGACATCACCGCGCC-----GCGCGCGCGCGCGCGCTCATCGA 4131886
Qy 240 pAlaAsnTyAsnSerIlys-----TyrAlaTyMetGlyThrSerMetAlaTh 257
Db 4131885 CCAGGAGTTCGGCGCAGAGCGCGGACCGGTACCTCACCATCTCTCCGTACGTGGCGAC 4131826
```

Percent Similarity: 43.33% Conservative: 56
Best Local Similarity: 28.97% Mismatches: 153
Query Match: 14.51% Indels: 68
DB: 14 Gaps: 14

US-09-985-689A-4 (1-433) x US-10-156-761-1 (1-9025608)

QY 8 ValiysAlaAspValAlaGlnAsnAsn-----Tyr 17
DB 919813 GTCCAGGCGGACATGCGCCGAGAGCAACGCGCAGATCGGTACGCGGGCGCGTGGGACGCC 6918872

QY 18 GlyLeuTyr-GlyGlnGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
DB 918873 GGGCTCAGCGGCGACGGCGTCACTGCGCGGTCTCGACACCGGCGTGCACACC----- 6918926

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysValThrAlaLeuTyrAlaLeuGly 57
DB 918927 -----ACTCACCCTCCGCGCGCGGTCTCCGCGAGCAAGCTTCATC 6918974

QY 58 ArgThrAsnAsnAlaSerProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 918975 GACGGGAGGAGTCCGCGCAGCGCAACGCGCAGCGACCGTCACTCCGCGCGTCCGC 6919034

QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 919035 GGGAGCGCGCGCCCTCCGCGGCGAGCGGCGCGGTCTCGCGCGGTCCAGCTCGCC 6919094

QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
DB 919095 GTCCGCAAGTGTCTAGCGACCGCGCGC-----GGAAGCGAGTCCCGATC 6919142

QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrp 129
DB 919143 ATCCGCGGATGGAATGGCGCGCGGAGCGTCTCGCGAGGATCTCTCGATGACCTC 6919202

QY 130 GlyAla---ProValAsnGlyAlaTyrThrAlaAsnSerArgGlnValAspGluTyrVal 148
DB 919203 GGATCGACGAGCGCAGCGGAGCGACCGCCATCGCGGAGCGCTCGACACCTCTCC 6919262

QY 149 ArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGlyProAsnSerGlyThr 168
DB 919263 GAGGAGACCGCGCGCCCTCTTCGTCGTCTCGCGCGGAAACACCGGTGCCCGCTCC-----TCG 6919319

QY 169 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnTyrArg 188
DB 919320 ATCGGTCGCGCGCGCGCGGACTCGCGCTGACGTCGCGCGC----- 6919364

QY 189 ProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSerArgGly 208
DB 919365 -----GTGACTCATCCGACCGCGCGCGCTACTTCAACGCGCGC 6919406

QY 209 AlaThrArg---AspGlyArgIleLysProAspValThrAlaProGlyThrPheIleLeu 227
DB 919407 CCGCGCCACGCGGCAACGCGCTCAAGCGGACTCGCGCAGCGCGGTTCGACATCCG 6919466

QY 228 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnTyrAsnSerLysTyr 247
DB 919467 GCCCGCGCTCCAGCTCGCGCGCGCGCGCTACTTCAACGCGCGC-----TAC 6919505

QY 248 AlaTyrMetGlyGlyThrSerMetAlaThrProfileValAlaGlyAsnValAlaGlnLeu 267
DB 919506 ACCTCCATGACGCTAGTCTGATCGCGACCGCGCGCTCGCGGGGTTCGCGCGCTCCTC 6919565

QY 268 ArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuIleLysAlaAla 287
DB 919566 GCCGAGCAGCACCCCGACTGACCGCGCGCGGTCTCAAGGACGCGCTGATGTCACGCTCC 6919625

QY 288 LeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProAsnGlyAspGlnGlyTrpGly 307
DB 919626 GAGCACTCGACGCTCCGTTATCATGTTGGGGCG-----GGT 6919664

QY 308 ArgValThrLeuAsnLysSerLeuAsnValAlaTyrValAsnGluAlaThrAlaLeuAla 327
DB 919664 -----GAGTTG 6919811

DB 919665 CGGGTCAGTGTGCGCGGACGCGCTCGCGCGC-----CGGCTCACCAGCGGCG 6919712

QY 328 ThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGlyLysProLeuLysIleSer 347
DB 919713 AGCGCGACCTCGCTTCCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919772

QY 348 LeuValTrpThrAspAlaProGlySerThrThrAlaSerTyrThrLeuValAsnAspLeu 367
DB 919773 GTCACTTACTTCAACTCTCCGACGCGCGCTC-----GAGTTG 6919811

QY 368 AspLeuValIleThr---AlaProAsnGly 376
DB 919812 AGCTCCCGGTGCGGGCGCGCGCGCGCGCT 6919841

RESULT 9
US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHISA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 2,21e-25 Length: 3417
Score: 325.00 Matches: 113
Percent Similarity: 45.56% Conservative: 46
Best Local Similarity: 32.38% Mismatches: 125
Query Match: 14.38% Indels: 67
DB: 14 Gaps: 12

US-09-985-689A-4 (1-433) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArg 37
DB 745 GGGTACGACGCGGAGCGGTCAAGATCGCGCTCTGGACACCGGTGTGAC----- 795

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
DB 796 -----GGACCCACCGCGACCTCAAGGACCGAGTGGCGGAGTCCCAAGACTTCTCC 846

QY 58 ArgThrAsnAsnAlaSerAspAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
DB 847 GCG 906

QY 78 Gly-----AsnAlaLeuAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
DB 907 GGCACCG 966

QY 93 PheGlnSerIleMetAspSerSerGly-----GlyLeu 103
DB 967 AACGCAAGGCTCTCGACGACACCGGCTCCGCGGACGACTCCCGGATCTCTCGCGCGGCAIG 1026

```

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OP INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 3624
Score: 113
Percent Similarity: 56
Best Local Similarity: 153
Query Match: 68
DB: 14 Gaps: 14

US-09-985-689A-4 (1-433) x US-10-156-761-5701 (1-3624)

Qy 8 ValLYeAlaAspValAlaGlnAsnAsh-----Tyx 17
Db 556 GTCGAGCGCGCATGGCGGAGAGAACGCCAGATCGTACGGGGGCCCGTGGGACGCC 615
Qy 18 GlyLeuTyxGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37

```

Db 1174 GGAATTCGCGCA-----CTTATCTTCAGCAAGCCCGCGGCACAGAACTT 1218
 Qy 282 SerLeuileLysAlaAlaLeuileAlaGlyAlaThrAspVal----- 295
 Db 1219 GATGAAGTCAAAACAGCTGCTAATGGACGGT---ACCGATTATGGAAAGATCCGATCCA 1275

Qy 296 -----GlyLeuGlyTyrProAsnGlyAspGln 304
 Db 1276 AATGTTATTCAGTGCAGGGTATCATCAACGCGAGAACNA 1311

RESULT 6

US-10-314-657-1
 ; Sequence 1, Application US/10314657
 ; Publication No. US2003017588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHEN, Ben
 ; APPLICANT: CHENG, Yi-Qiang
 ; APPLICANT: TANG, Gong-Li
 ; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
 ; FILE OF INVENTION: Syntheses and Methods of Use
 ; FILE REFERENCE: 054030-0021
 ; CURRENT APPLICATION NUMBER: US/10/314,657
 ; CURRENT FILING DATE: 2002-12-09
 ; PRIOR APPLICATION NUMBER: PCT/US02/08937
 ; PRIOR FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: US 60/278,935
 ; PRIOR FILING DATE: 2001-03-26
 ; NUMBER OF SEQ ID NOS: 214
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 1
 ; LENGTH: 135638
 ; TYPE: DNA
 ; ORGANISM: Streptomyces atroolivaceus
 US-10-314-657-1

Alignment Scores:

Pred. No.: 4,22e-24 Length: 135638
 Score: 334.50 Matches: 137
 Percent Similarity: 42.17% Conservative: 57
 Best Local Similarity: 29.78% Mismatches: 155
 Query Match: 14.80% Indels: 111
 DB: 14 Gaps: 23

US-09-985-689A-4 (1-433) x US-10-314-657-1 (1-135638)

Qy 6 GlyileValLysAlaAspValAlaGlnAsnAsn----- 16
 Db 8067 GCGAGGGTGAAGCGATCTCGCGCACTCCACCCCGCAGATCGCGCGCAGAGGTATGG 8126
 Qy 17 ---TyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAspThrGlyLeuAspThr 35
 Db 8127 GCGAGGGCGACACCGGCGCAGGAGGTGCGATGCTCGACAGCGCGCGACACC 8186
 Qy 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysile-----ThrAlaLeu 53
 Db 8187 -----GACACCCGCACTGTGCGGCGAGGTGTCGACAGCGCCACG 8228
 Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAla 73
 Db 8229 TTCGTCCTCCGCGCAG---GACGACATCGCGCACTACACGCCACGCGCAGCGACGTCGCC 8285
 Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88
 Db 8286 TCGACCATCTCGGCGACCGGCGAGCGCCCTCCGACGGCAAGGCGGGTGTGCGCTCCGCG 8345
 Qy 89 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyLeuGlyLeuProSer 108
 Db 8346 GCGCGGCTGTCCGTCGCGCAAGGTGTCAACTCCGAGGCG-----AGCGGCGAGGAATCG 8399
 Qy 109 AsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHis 125
 Db 8400 TGGATC-----ATCGCGGGATGGAGTGGGCGCGCCCGCAGACGAGGCGGAGGATCATC 8453

RESULT 7

US-10-156-761-5701
 ; Sequence 5701, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN

Qy 126 ThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAlaAsn-----SerArg 142
 Db 8454 AGCATGAGCCTGGGC-----GGCGCGGTGACAAAGAACGACCCCGATCAGCGAC 8501
 Qy 143 GlnValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGlu 162
 Db 8502 GCGGTCGACGAACCTCAGCCACGACGCGCGCGTGTTCGTGATCGCGCGGCAACGCG 8561
 Qy 163 GlyProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 8562 GCGCGCACTCC-----ATCAGCAGCGCGGTGCGGCAGACTCCGCGCTCAGCGTGGC 8615
 Qy 183 AlaThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAla 202
 Db 8616 GCC-----GTGACTCCACCGACCGCTCGCC 8642
 Qy 203 GlnPheSerSerArgGly---AlaThrArgAspGlyArgIleLysProAspValThrAla 221
 Db 8643 GACTTCTCCAGCAGGCGCGCGTGCAGCGGCGCGGTGAAGCGGAGATCACCAGCG 8702
 Qy 222 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 241
 Db 8703 CCGCGCGTCGACATGCTCGCGCGCGTTCGCACTACAAGCGCGGTCTCGGCTAC----- 8756
 Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAla 261
 Db 8757 -----TACACCGATGAGCGGCGCGTGCATGGCGCGCGCGCGCGCGCGCGCG 8801
 Qy 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
 Db 8802 GGTGTGCGCGCGCTCTCGCGCGCGAGCACCGCGCGCGCGCGCGCGCGCGCGCGCG 8861
 Qy 282 SerLeuIleLysAlaAlaLeuIleAlaGlyAlaThrAspValGlyLeuGlyTyrProAsn 301
 Db 8862 GCACCTGGTCAGCAGCGCGCAAGCAACGCGCGG-----TACACCGCG 8903
 Qy 302 GlyAspGlnGlyTyrGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyrValAsn 321
 Db 8904 TACCAGCGCGCGCGCGCGCGTTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8963
 Qy 322 GluAlaThrAlaLeuAlaThrGlyGlnLysAlaThrTyrSerPheGlnAlaGlnAlaGly 341
 Db 8964 ACCAGACCGCGCTACTCC-----GGCTTCCAC----- 8990
 Qy 342 LysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr-----Thr 358
 Db 8991 -----ACGTGCG 9029
 Qy 359 AlaSerTyrThrLeuValAsnAsp-----LeuAspLeuValIle-----Thr 372
 Db 9030 GTGAGGTACCAACAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9089
 Qy 373 AlaProAsnGlyGlnLysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrp 392
 Db 9090 GTCCCGCGCGGTG-----TTCAGCCTCTCCGAGGACCAT----- 9125
 Qy 393 AspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrlle 412
 Db 9126 -----GTACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9155
 Qy 413 IleGluVal---GlnAlaTyrAsnValProSerGlyProGlnArgPheSerLeuAlaIle 431
 Db 9156 GTCACCTCAGCGCGCTCTGACAAAGCTGGCGGCGCGCGCGCGCGCGCGCGCGCG 9215


```
Db 913 TATCAATCGGTCTCCAGCAGCTGCAAGCAAGTTATTACAGTTGGAGCCGTTCAACAAG 972
Qy 187 TyArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 973 TAT-----GATGTTTAAACAAGCTTCTCAAGC 999
Qy 207 ArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGlyThrPheIle 226
Db 1000 AGAGGCCCAACTGCAGACGGCAGGCTTAAGCTGAGTTGTTGCTCCAGGAACCTGGATA 1059
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerPheThrAlaAsnTyrAsnSerLys 246
Db 1060 ATTGCTGCCAGCAAGT-----GGAACTAGCATGGTCAACCAATTAATGACTAT 1110
Qy 247 TyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 1111 TACACAGCAGCTCTCGGACATCAATGGCAACTCTCACTAGCTGGTATTCACGCCCTC 1170
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrPro-----LysProSerLeu 283
Db 1171 TTGCTCAA-----GCACACCCGAGCTGGACTCCAGACAAAGTAAACACGCCCTC 1221
Qy 284 IleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly---LeuGlyTyrProAsnGly 302
Db 1222 ATAGAAACTGCTGATATCGTAAAGCCAGATGAATAGCCGATATAGCCTAC----- 1272
Qy 303 AspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaLysValAlaGln 322
Db 1273 -----GGTGCAGGTAGGTTAATGATCAAGGCTATAAAC-----TACGATAACTAT 1320
Qy 323 AlaThrAlaLeuAlaThrGlyGlnLysAla-----ThrTyrSerPheGln 337
Db 1321 GCAAGCTAGTGTCTACTGGATATGTTGCCAACAAAGGCGAGCCAACTCACCAGTTGGT 1380
Qy 338 AlaGlnAlaGlyLysProLysIleSerLeuValThrAspAlaProGlySerThr 357
Db 1381 ATTAGCGGAGCTTCGTTGTAAGTGCACATTATAGTGGCAATGCCAAT----- 1431
Qy 358 ThrAlaSerTyrThrLeuValAsnAspLeuAspValIleThrAlaProAsnGlyGln 377
Db 1432 -----ACGACCTTGATCTTTACTCTACGATCCCAATGGAAC 1470
Qy 378 LysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsn 397
Db 1471 CAGGTT-----GACTACTCTTAC-----ACCGCTACTATGGA 1503
Qy 398 ValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrIleIleGluValGlnAla 417
Db 1504 TTCGAAAGGTTGGTTATTACACCCCACTGATGGACATGGACAATTAAAGTTGTAAAC 1563
Qy 418 TyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432
Db 1564 TAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGTA 1599
```

RESULT 5

```
US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934
Alignment Scores:
Pred. No.: 9,3e-28 Length: 1329
Score: 341.00 Matches: 106
Percent Similarity: 44.58% Conservative: 42
Best Local Similarity: 31.93% Mismatches: 116
Query Match: 15.09% Indels: 68
DB: Gaps: 12
US-09-985-689A-4 (1-433) x US-09-974-300-1934 (1-1329)
Qy 11 AspValAlaGlnAsnAsnTyrGlyLeuTyrGlyGlnGlnValValAlaValAlaAsp 30
Db 406 GAAGTGTGTGCAACAATCAGACGCTGACAGGCAAGAGAGTACAGTCGCTGTCATTGAT 465
Qy 31 ThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIle 50
Db 466 ACGGCGTA-----TACCTCAGCAAGATCTTGAAGCAGATC 504
Qy 51 ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGly 69
Db 505 AGGCTTTTCAAGACTTTATCAACAGAGAACAGAACCTTATGATGACAAATGGCAGCGC 564
Qy 70 ThrHisValAlaGlySerValLeuGlyAsnAlaLeu-----AsnLysGly 84
Db 565 ACACACTGCGCGGTGATGCTTGGGAACAGGAGCGGCTCATCGGGTCAGTACCGCGGA 624
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerSerGlyLeuGly 104
Db 625 CCTGCTCTGAAGCAGAACTTGTGGTGTAAAGTATTGGACAAATGGGATCC---GGA 681
Qy 105 GlyLeu-----Pro 107
Db 682 TCGTCTGAAACCCGTCATTCAGGCGTAGATTGGTCATTCAATTCAATAAGGAAATTCCT 741
Qy 108 SerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThrAsn 127
Db 742 GATGATCCGATCGACATTATTTCATGCTGCTGGTGCAGAGCCCTTGGCTACGAGAAT 801
Qy 128 SerTrpGlyAlaPro-----ValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
Db 802 GAAGAAGAAGATCCAGTCGTTAAAGCTGTTTCATGACGATGGGACGAGGC----- 852
Qy 144 ValAspGluTyrValArgAsnAsnAspMetThrValLeuPheAlaAlaGlyAsnGluGly 163
Db 853 -----ATTGTTGATGTGGCGAGCGCGCACTCCCGT 885
Qy 164 ProAsnSerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 886 CCGTATGCGCAACGATTCAGCCGCGGTGTCCAGCAGCAAGATATTATACAGTCGAGCC 945
Qy 184 ThrGluAsnTyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGln 203
Db 946 TTGATGACAG-----GATACAGTCAGCGGAGGATGACGATGTCGCTCT 993
Qy 204 PheSerSerArgGlyAlaThrArgAspGlyArgIleLysProAspValThrAlaProGly 223
Db 994 TATTCAAGCAGAGGCGCGCAATCTATGTTCAAGTCAACCGGACTTGTGTTGTTACCGGCG 1053
Qy 224 ThrPheIleLeuSerAlaArgSer-----SerLeuAlaProAspSerSerPheTrpAla 241
Db 1054 ACAATATTACGTCGCTTCGTTTCCCGGATCTTTTCTCGATAAGCTGCAAAAAACAAC 1113
Qy 242 AsnTyrAsnSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAla 261
Db 1114 AGAGTCGGCACAAATATATGACATTGTCGGAACCTCGATGGTGTAGCCCATCTGGCA 1173
Qy 262 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 281
```


Db 217 TCATACGACCTGGTACTGGAGCGAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGCA 276
Qy 89 AlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGlyLeuGlyLeu 106
Db 277 GCTAAGCTGGCGGAATTAAGTTCTAGGTGCGGATGGTTCTGGAAGCATATCTACTATA 336
Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThr 126
Db 337 ATTAAGGAGTTGAGTGGCGGCTGTGATAACAAGATTAAGTACGAATTAAGTTCATTAT 396
Qy 127 AsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgGlnValAspGlu 146
Db 397 CTTTCTCTGTTTCAAGCCAGAGCTCAGATGCTACTCAGCTCTAAGTCAGGCTGTTAAT 456
Qy 147 TyrValArgAsnAspMetThrValLeuPheAlaGlyAsnGluGlyProAsnSer 166
Db 457 GCACGGTGGGATGCTGATGATTGTTGTTGTCGCCCTGGAACAGTGGACCTTAACAG 516
Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186
Db 517 TATACAAATCGTTCTCCAGCAGCTGCAAGCAAGTTATTACAGTTGGAGCGCTTGACAAG 576
Qy 187 TyrArgProSerPheGlySerIleAlaAspAsnProAsnHisIleAlaGlnPheSerSer 206
Db 577 TAT-----GATGTTATAACAAGCTTCTCAAGC 603
Qy 207 ArgGlyAlaThrArgAspGlyValIleLysProAspValThrAlaProGlyThrPheIle 226
Db 604 AGAGGCCCACTGCAGACGGCAGCTTAAGCTGAGGTTGTTGCTCCAGGAAGTGGATA 663
Qy 227 LeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnThrAsnSerLys 246
Db 664 ATTGTCGCAGAGCAAGT-----GGAAGTACATGGGTCAACCAATTAATGACTAT 714
Qy 247 TyrAlaTrpMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGln 266
Db 715 TACACAGCAGCTCTCTGGACATCAATGGCACTCTCAGTAGCTGTATGAGCCCTC 774
Qy 267 LeuArgGluHisPheIleLysAsnArgGlyIleThrPro-----LysProSerLeu 283
Db 775 TTGCTCCAA-----GCACACCGAGCTGGACTCCAGACAAAGTAAACAGCCCTC 825
Qy 284 IleLysAlaAlaLeuIleAlaGlyAlaThrAspValGly--LeuGlyTyrProAsnGly 302
Db 826 ATAGAACTGCTGATGTAATGTAAGCAGATGAATAGCGATATAGCTTAC-----876
Qy 303 AspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTrpValAsnGlu 322
Db 877 -----GGTCAGGTAGGTTAATGCATACAGGCTATAAC-----TACGATAACTAT 924
Qy 323 AlaThrAlaLeuAlaThrGlyGlnLysAla-----ThrTyrSerPheGln 337
Db 925 GCAAGAGTAGTGTTCACATGGATATGTTGCCAACAAAGCAGCCAACTCACCAGTTCGTT 984
Qy 338 AlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpThrAspAlaProGlySerThr 357
Db 985 ATTAGCGAGCTTCGTTGCTAAGTCCACATTTATCTACGGCAATGCCAAT-----1035
Qy 358 ThrAlaSerTyrThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyGln 377
Db 1036 -----AGCAGCTTGTATCTTACCTCTACGATCCCAATGGAAC 1074
Qy 378 LysTyrValGlyAsnAspPheSerTyrProTyrAspAsnAsnTrpAspGlyArgAsnAsn 397
Db 1075 CAGGTT-----GACTACTCTTAC-----ACCGCTACTATGGA 1107
Qy 398 ValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrIleIleGluValGlnAla 417
Db 1108 TTCGAAAAGGTTGTTATTACAAACCACTGATGGAACATGGAACAATTAAGTTGTAAGC 1167
Qy 418 TyrAsnValProSerGlyProGlnArgPheSerLeuAlaIleVal 432

Db 1168 TAC-----AGCGGAAGTGCAAACTATCAAGTAGATGTGGTA 1203

RESULT 4

US-10-090-624-15
; sequence 15, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/10/090,624

; CURRENT FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: 09/445,472

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997

; PRIOR FILING DATE: 1997-06-10

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 15

; LENGTH: 1962

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic

US-10-090-624-15

Alignment Scores:

Pred. No.:	6,82e-36	Length:	1962
Score:	416.50	Matches:	137
Percent Similarity:	44.14%	Conservative:	55
Best Local Similarity:	31.49%	Mismatches:	172
Query Match:	18.43%	Indels:	71
DB:	13	Gaps:	17

US-09-985-689A-4 (1-433) x US-10-090-624-15 (1-1962)

Qy 16 AsnTyrGlyLeuTyrGlyGlnGlnValValAlaValAlaAspThrGlyLeuAspThr 35

Db 454 AACTTGGGATATGATGTTCTGGAATCAATAGGAATAATTGACACTGGAATTGAC--- 510

Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeu---- 53

Db 511 -----GCTTCTATCCAGATCTCCAGGAAAAGTAATTTGGTGGGTAGAT 555

Qy 54 TyrAlaLeuGlyArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAla 73

Db 556 TTTGTCAATGGTAGG---AGTTATCCATACGATGACCATGACATGGACATGGAACCTCATGTAGCT 612

Qy 74 GlySerValLeuGly-----AsnAlaLeuAsnLysGlyMetAlaProGln 88

Db 613 TCAATAGCAGCTGGTACTGGAGCACAAGTAATGGCAAGTACAGGGAATGGCTCCAGCA 672

Qy 89 AlaAsnLeuValPheGlnSerIleMet-----AspSerSerGlyGlyLeuGlyLeu 106

Db 673 GCTAAGCTGGCGGAATTAAGTTCTAGGTGCGGATGTTCTTGGGAAGCATATCTACTATA 732

Qy 107 ProSerAsnLeuAsnThrLeuPheSerGlnAlaTrpAsnAlaGlyAlaArgIleHisThr 126

Db 733 ATTAAGGAGGTTGAGTGGGCGCGTGTATACAAAGATAAGTACGGAATTAAGGTCAATTAAT 792

Qy 127 AsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAlaAsnSerArgGlnValAspGlu 146

Db 793 CTTTCTCTGTTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAAT 852

Qy 147 TyrValArgAsnAsnAspMetThrValLeuPheAlaIleGlyAsnGlyGlyProAsnSer 166

Db 853 GCACGGTGGGATGCTCGGATTAAGTTCTTGTGTTGCTGCGCTGGAACAGTGGACCTAACAG 912

Qy 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-4
Perfect score: 2260
Sequence: 1 NDVARGIVKADVAQNNGLY.....EVQAVNPSPGQRFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q=/cn2_1/USPTO_spool/US09985689/runat_31032004_161809_4271/app_query.fasta_1.3498
-DB=Published_Applications_NA -QWTF=fastap -SUFFX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -PAIRX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09985689 @CGN 1.1 601 @runat_31032004_161809_4271
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cn2_6/ptodata/1/pubpna/PTC_NEW_PUB.seq:
3: /cn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:
4: /cn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cn2_6/ptodata/1/pubpna/PTCUS_PUBCOMB.seq:
7: /cn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17: /cn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18: /cn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description

1 1987.5 87.9 1305 15 US-10-385-662-1
2 438 19.4 1377 13 US-10-090-624-11
3 416.5 18.4 1236 13 US-10-090-624-2
4 416.5 18.4 1362 13 US-10-090-624-15
5 341 15.1 1329 9 US-09-974-300-1934
6 334.5 14.8 135638 14 US-10-314-657-1
7 328 14.5 3624 14 US-10-156-761-5701
8 328 14.5 9025608 14 US-10-156-761-1
9 325 14.4 3417 14 US-10-156-761-1
10 325 14.4 9025608 14 US-10-156-761-1
11 313 13.8 4765 13 US-10-090-624-5
12 300.5 13.3 3743 10 US-09-927-827-29
13 298 13.2 1560 15 US-10-084-846A-113
14 298 13.2 59816 15 US-10-084-846A-1
15 298 13.2 59816 15 US-10-084-846A-2
16 277.5 12.3 3303 14 US-10-156-761-5384
17 277 12.3 3788 10 US-09-927-827-33
18 276.5 12.2 2166 12 US-10-344-231-17
19 276.5 12.2 2166 12 US-10-363-332A-17
20 270 11.9 1140 8 US-08-322-678-11
21 270 11.9 1140 16 US-10-323-324-11
22 270 11.9 1143 14 US-10-313-853-6
23 270 11.9 2588 12 US-10-344-231-20
24 270 11.9 2588 12 US-10-363-332A-20
25 269 11.9 1140 8 US-08-322-678-12
26 269 11.9 1140 16 US-10-323-324-12
27 263 11.6 840 14 US-10-203-812-1
28 261.5 11.6 1306 9 US-09-966-921A-1
29 261.5 11.6 1330 9 US-09-966-921A-5
30 260.5 11.5 1485 9 US-09-974-300-1938
31 259.5 11.5 2287 12 US-10-344-231-18
32 259.5 11.5 2287 12 US-10-363-332A-18
33 257.5 11.4 1971 9 US-09-974-300-1935
34 248 11.0 3884 10 US-09-927-827-34
35 247.5 11.0 3452 10 US-09-927-827-30
36 245.5 10.9 2192 12 US-10-424-599-112429
37 243 10.8 1140 15 US-10-146-905A-9
38 239 10.6 1140 9 US-09-920-118-13
39 232 10.3 1332 14 US-10-156-761-5689
40 231 10.2 1575 12 US-10-344-231-19
41 231 10.2 1575 12 US-10-363-332A-19
42 228 10.1 1457 14 US-10-202-339-1
43 227.5 10.1 522 10 US-09-824-893A-98
44 227 10.0 1497 8 US-08-322-678-6
45 227 10.0 1497 9 US-09-960-854B-1

ALIGNMENTS

RESULT 1

US-10-385-662-1

; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NORUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSURUMI
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938U0
; CURRENT APPLICATION NUMBER: US/10385.662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

Search completed: April 4, 2004, 12:02:30
Job time : 90.4512 secs

```
QY 204 PheSerSerArgGly---AlaThrArgAspGlyAlaGlyLeuValProAspValThrAlaPro 222
Db 1508 TTCTCTCTCACGGCCCGCTCGGACGCGGCGCATCAAGCCGAGTTCACGGCTCCC 1567
QY 223 GlyThrPheLeuSerAlaArgSerLeuAlaProAspSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACGCGCCCTCGCGGAGGCAACAGCATCGGCCAGGAGTTCGGTGAG 1627
QY 243 TyrAsnSerLysTyAlaTrpMetGlyGlyThrSerMetAlaThrProileValAlaGly 262
Db 1628 GGACGGCGCGGTATGACCATCTCCGGCACGTGATGGGACCCCGACAGTTCGGCGGC 1687
QY 263 AsnValAlaGlnLeuArgGlyHisPheLeuLysAsnArgIlyleThrProLysProSer 282
Db 1688 GCGGGCGGCTCTGAAGCAGCAG-----CACCCCGAC 1720
QY 283 LeuLeuLysAlaAlaLeu-----lleAlaGlyAlaThrAspValGlyLeuGlyTyr 299
Db 1721 TGGACCTCGCGCAACTGAAGGCGCGCTCACCGCTCCACAGGGCGGC---AAGTAC 1777
QY 300 ProAsnGlyAspGlnGlyTrpGlyArgValThrLeuAsnLysSerLeuAsnValAlaTyr 319
Db 1778 ACCCGTTCAGCAGGTTTCGGCGGATCCAGGCGGACAGGCGCTCCAGCAGACCGTG 1837
QY 320 ValAsnGluAlaThrAlaLeuAlaThrGly----- 329
Db 1838 ATCGCGCAGCCGGTCTCGGTGAGCTTCGGCGTCCAGCAGTGCCTCCACACCGACGACGAG 1897
QY 330 -----GlnLysAlaThrTrpSerPheGlnAlaGlnAlaGlyLysProLeuLysIle 346
Db 1898 CCGGTACCAGCAGCTGACCTACCGCAACCTCGGCACCCAGGAGCTCACGCTGAAGCTG 1957
QY 347 SerLeuValTrpThrAspAlaProGlySerThrThrAlaSer-----TyrThrLeuVal 364
Db 1958 AGCTGACCGCCACCCAGGCGGCAAGGCGGCGCGCGGCTTCTTCAGCTGGGC 2017
QY 365 AsnAspLeuAspLeuValIleThrAlaProAsnGlyGlnLysTyrValGlyAsnAspPhe 384
Db 2018 -----GCCACACGGTGACCGTCCCGGGGCGGC----- 402
QY 385 SerTyProTyAspAsnAsnTrpAspGlyArg-----AsnAsnValGlu 399
Db 2048 AGCGCTCGGTGCATGACCGCGCACACCGCGCTCGGGCGCACGCTGACGCGCGTAC 2107
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrIle----- 412
Db 2108 TCGCGTACGTGTGTCGCCACGGCGGCGGCGGCGGACGTCGCCGACGCGCGCGCGTGCAG 2167
QY 413 IleGluValGlnAlaTyrAsnVal-ProSerGly 423
Db 2168 CCGGAGTTCAGTGTGACGACGTGACCGTCCGCGC 2201

RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
;
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
```

```
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICANT: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 3 41e-24 Length: 2809
Score: 340.50 Matches: 131
Percent Similarity: 40.93% Conservative: 54
Best Local Similarity: 28.98% Mismatches: 171
Query Match: 15.07% Indels: 96
Gaps: 18
US-09-985-689A-4 (1-433) x US-09-514-340-1 (1-2809)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCAAGGCGCGTGAAGATCGCGTCTCGACACCGGTGTGACACGAGC--- 1054
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTGAAGGCGGCGGTACCGGTCACAGAACTTCACC 1099
QY 58 ArgThrAsnAsnAlaSerAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCCGCGCGCGCGCGCGCAAGGTGGCGCACCGCACCCACCGCTCGATCGCGCGG 1159
QY 78 GlyAsnAlaLeuAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGACGGGCGGCCAGTCCAAAGGCAAGTACAGGGCGTCCACCGCGCGCGGATCCTC 1219
QY 93 PheGlnSerIleMetAspSerSerGlyGlyLeuGlyLeuProSerAsnLeuAsnThr 112
Db 1220 AACGGCAAGGTCTCTCGACGACTCC-----GGTTTCGCGCAGCACTCCGGCATC 1267
QY 113 LeuPheSerGlnAlaTrpAsnAla-----GlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1268 CTCGCCGCGATGGAGTGGCGGCGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGC 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrAlaAsnSerArgGln 143
```

```

Db 1595 GAGGCAACACATCGCGCAGAGAGTCTGGTGAGGACCGCGCGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCTGATGGCCACCCCGCACCTCGGGGGCGGGCGCTCTCTGAAGCAGCAG--- 1711
Qy 273 IleIysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACGCCGACTCGGACCTCGCGCGAACTGAAGGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACCGGCTCCACCAAG---GGCGGCAAGTACACCCCGTTGAGCAGGGTTCGGGCCCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeuSer--- 328
Db 1805 ATCCAGGCGCACAGGGCGTCCAGCAGACCGGTATCGCGCACCGGCTCTCGGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGGTCCAGAGTGGCCGACACCGACGACGCGCGTCCACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTyrSerAsp----- 353
Db 1925 AACCTCGGCACCCAGGACGTACGCTGAGCTGACGTGACGCGCACCGCCACCGCCCAAGGGC 1984
Qy 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGCGCGCGCGGGCTTCTTACGCTGGCGGCCACCCACG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGCGCGCGCGCAGC---GCCTCCGTGACATGACCGCGC----- 2071
Qy 390 AspAsnAsnTyrAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCGCGCTCGCGGCGCACGCTGGACGCGCGGTACTCGCGGTACGCTGCTGCC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGCGGCGACACGCTCCGCGCGCGCGCGGTGTCAGCGCGAGGTGAGTCGTACGAC 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCCGCGCACATCGGCC 2210

```

Search completed: April 4, 2004, 01:12:51
Job time : 364.684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.5977 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-6
Perfect score: 2250
Sequence: 1 NDVARGIVKADVAGSSVGLY.....EVQAYNVQGPQAFSLAIVN 434

Scoring table:

BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2_1/USFTO_spool/US0995689/runat_31032004_161807_4184/app_query.fasta_1.3498

-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=humam40.cdi

-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NORSE=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000

-USER=US0995689 @CGN 1 1 142 @runat_31032004_161807_4184 -NCPU=6 -ICPU=3

-NO WMAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6

-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*

3: /cgn2_6/ptodata/2/ina/5A COMB.seq:*

4: /cgn2_6/ptodata/2/ina/5B COMB.seq:*

5: /cgn2_6/ptodata/2/ina/pctus COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2143	95.2	1923	4	US-09-509-814A-5
2	2141	95.2	1923	4	US-09-509-814A-7
3	2122	94.3	1920	4	US-09-509-814A-3
4	2116.5	94.1	3003	2	US-08-873-479-41
5	450.5	20.0	1977	3	US-08-894-818B-2
6	450.5	20.0	1977	4	US-09-445-472-11
7	406.5	18.1	1236	4	US-09-445-472-2
8	406.5	18.1	1566	3	US-08-894-818B-4
9	406.5	18.1	1962	3	US-08-894-818B-34
10	406.5	18.1	1962	4	US-09-445-472-15
11	405.5	18.0	1977	3	US-08-894-818B-6
12	360.5	16.0	2539	3	US-09-000-016-3

13	360.5	16.0	2539	4	US-09-514-340-3	Sequence 3, Appli
14	355.5	15.8	2809	3	US-09-000-016-1	Sequence 1, Appli
15	355.5	15.8	2809	4	US-09-514-340-1	Sequence 1, Appli
16	318.5	14.2	2532	1	US-07-671-376C-4	Sequence 4, Appli
17	317	14.1	2835	1	US-08-750-532-2	Sequence 2, Appli
18	317	14.1	4765	1	US-08-750-532-8	Sequence 8, Appli
19	317	14.1	4765	3	US-08-894-818B-7	Sequence 7, Appli
20	317	14.1	4765	4	US-09-445-472-5	Sequence 5, Appli
21	305.5	13.6	1859	3	US-08-894-818B-15	Sequence 15, Appli
22	290	12.9	898	1	US-08-750-532-7	Sequence 7, Appli
23	286.5	12.7	564	1	US-08-750-532-11	Sequence 11, Appli
24	286.5	12.7	564	3	US-08-894-818B-14	Sequence 14, Appli
25	272	12.1	1306	4	US-09-966-921A-1	Sequence 1, Appli
26	272	12.1	1330	4	US-09-966-921A-5	Sequence 5, Appli
27	267	11.9	807	1	US-07-706-691G-56	Sequence 56, Appli
28	267	11.9	807	1	US-08-254-021-56	Sequence 56, Appli
29	267	11.9	807	2	US-08-618-446-56	Sequence 56, Appli
30	267	11.9	807	3	US-08-980-135-56	Sequence 56, Appli
31	267	11.9	807	4	US-09-585-798-56	Sequence 56, Appli
32	266.5	11.8	10216	2	US-08-875-154-1	Sequence 1, Appli
33	266	11.8	807	1	US-08-566-369-2	Sequence 2, Appli
34	266	11.8	807	3	US-09-074-331-2	Sequence 2, Appli
35	266	11.8	807	5	PCT-US95-01937-2	Sequence 2, Appli
36	265	11.8	807	1	US-07-706-691G-57	Sequence 57, Appli
37	265	11.8	807	1	US-08-254-021-57	Sequence 57, Appli
38	265	11.8	807	2	US-08-618-446-57	Sequence 57, Appli
39	265	11.8	807	2	US-08-980-135-57	Sequence 57, Appli
40	265	11.8	807	4	US-09-585-798-57	Sequence 57, Appli
41	264	11.7	840	1	US-08-434-255-7	Sequence 7, Appli
42	264	11.7	840	1	US-08-459-967-7	Sequence 7, Appli
43	264	11.7	840	1	US-08-460-327-7	Sequence 7, Appli
44	264	11.7	840	1	US-08-459-871-7	Sequence 7, Appli
45	264	11.7	840	3	US-09-024-532-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, NASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores: 1.53e-213 Length: 1923
Pred. No.: 2143.00 Matches: 406
Score: 2143.00
Percent Similarity: 97.93% Conservative: 19

```
Best Local Similarity: 93.55% Mismatches: 9
Query Match: 95.24% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-6 (1-434) x US-09-509-814A-5 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValValValValAlaGlnSerSerGlyLeuTyr 20
Db 619 AATGATGTTGGCGTGGATTTGCAAGAGCGATGTGGCTCAGAGCAGCTACGGTGTAT 678
Qy 21 GlyGlnGlyValValValValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGAAGAAGACAGATCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleThrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGCATGAGCCCTTCGCGGGAATTTACTGCAATTTATGCAATTTGGGAGCGAAT 798
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGCTATGCTAGCATGCTGGCTGGCTGGCTGGCTGGCTGGC 858
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAACTCATGATGATAGC 918
Qy 101 AsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 919 GGTGGGGAGCTTGGAGGACTACCTTCCAACTCTCAAACTTATTCAGCCAAAGCATACAT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaProValAsnGlyAlaTyrThr 140
Db 979 GCTGGTCCAGAACTTCATACAACTCTCGGGAGCAGCATGTAATGGGCTTACAAACA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspValAlaValLeuPheAlaAla 160
Db 1039 GATTCAGAAATGTGGATGATCTGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGGATGAGAGACCGACCGCGGACCATCAGTGCACCGACGACAGCTAAAGATGCAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCAAGCTTTGGGTCTTATGCGCAATATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCAGATGCTCTTTCAGTGGACCGACAAAGATGACGATCAAAACCGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1279 ATGGCACCGGGAACGTTTCATATCATCTATCAGCAAGATCTTCTTGCACCGGATCTCT 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATGACAGTAAATATGATACATGTTGGTGAACGTCCTTCCGCTACCGCATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
Db 1399 GTTGTGGAACGTTGGCAGACCTTCGTGAGCATTTTGTGAAAAACAGAGGCAACACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaThrAspIleGlyLeuGlyTyr 300
Db 1459 AAGCCTTCTCTATTAAAAAGCGCACTGATTCGCGTGCAGTGCATCGGCTTGGCTAC 1518
Qy 301 ProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 1519 CCGAACGTTACCAAGGATGGGACGATGACATGGATTAATCCCTGAAGCTTGCCTAT 1578
Qy 321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
Db 1579 GTGAACGATGCTCCTCTATCCACCACGCCAAAGCGAGCTACTCGTTTACTGCTACT 1638
```

```
341 SerGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrSerAla 360
1639 GCCGCAAGCCTTTGAAAATCTCCCTGGTATGCTCTGATGCCCTCGGACCAACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCTATCCGCTCCAAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 400
1759 GTAGGAATGACITTTACTTCGCCATACAATGATAACTGGGATGGCGCAATAACGTAGAA 1818
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
1819 AATGATTTTATTAATGACCAACAAAGCGGACGTATACAAATTCAGGTACAGGCTTATAAC 1878
Qy 421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
1879 GTACCGGTGGACCACAGACCTTCTCGTTGGCAATTTGTAAT 1920

RESULT 2
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKKI, KATSUHIKA
; APPLICANT: KUSOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0632-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-7

Alignment Scores:
Pred. No.: 2,48e-213 Length: 1923
Score: 2141.00 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 9
Query Match: 95.16% Indels: 0
DB: Gaps: 0

US-09-985-689A-6 (1-434) x US-09-509-814A-7 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValValValAlaGlnSerSerGlyLeuTyr 20
Db 619 AATGATGTTGGCGTGGATTTGCAAGAGCGATGTGGCTCAGAGCAGCTACGGTGTAT 678
Qy 21 GlyGlnGlyValValValValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGAAGAAGACAGATCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleThrAlaLeuGlyArgThrAsn 60
```

Db 739 TCGATGCATGAAGCCCTCCGGGGAATAATCTGCTATTATATGCTATTGGGACGCAAT 798
Qy 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGGTCATGGTACGCAATGGCTGGCTCGTATTAGGAACGGC 858
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCATCTCATGATGATAGC 918
Qy 101 AsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 919 GGTGGGGACTTGGAGGACTACTCTCGAATCTGCAAACTTATTTCAGGCAAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 140
Db 979 GCTGTGTCGAGAAATCATACAAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACACA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAla 160
Db 1039 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
Qy 161 GlyAsnGlyLeuProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 1099 GCGAATGAAGGACCGAAGCGGGAACCATCAGTCACCGAGGCACAGCTAAATAATCAATA 1158
Qy 181 ThrValGlyAlaThrGlnAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCAAGCTTTGGGTCTTATGCGGACAAATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCAGCTTCTCTTCACGTGGACCGCACAAAGGATGGACGATCAACCGGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGAGCTTCTACTATCAGCAAGATCTTCTTCGACCGGATTCCTCCCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGGCGAACCATGACAGTAATAATGATATGATCATGCTGGTGGAACTCCATGCTACACCGATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
Db 1399 GTTGCTGGAAACGTGGCAGCTTCGTGAGCATTTTGTGAAACACAGAGGCATCACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr 300
Db 1459 AAGCCTTCTCTATTAAGCGGCACCTGATTGCCGHTGCAGCTGACATCGGCCTTGGCTAC 1518
Qy 301 ProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTCGATAAATCCCTGAACGTTGCCTAT 1578
Qy 321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
Db 1579 GTGAACGAGTCCAGTCTCTATCCACGACGCAAAAGCGACGCTACTCGTTTACTGCTACT 1638
Qy 341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
Db 1639 GCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGCTCCCTCGAGCACAATGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
Db 1699 TCGGTAACGCTTGTCAATGATCTGGACCTTGTCAATACCGCTCCAAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1759 GTAGGAATGACCTTACTTCGCCATACATGATTAATCTGGATGGCGCGCAATACGTAGAA 1818
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
Db 1819 AATGTATTTATTAATGACCACAAAGCGGACGCTATACAAATGAGTACAGGCTTATAAC 1878

Qy 421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTTGGACCACAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1920)
US-09-509-814A-3

Alignment Scores:
Pred. No.: 2,37e-211 Length: 1920
Score: 2122.00 Matches: 400
Percent Similarity: 97.70% Conservative: 24
Best Local Similarity: 92.17% Mismatches: 10
Query Match: 94.31% Indels: 0
Gaps: 4

US-09-985-689A-6 (1-434) x US-09-509-814A-3 (1-1920)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGATATTGTCAAGCGGATGTGGCACAGACAGCTACGGTTTGAT 675
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGGCCAGATGTGCGAGTTGCGGATACTGGATTGGATACAGGAAGAACGACGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGATGAAGCCTTCCGCGGTAATAAACACACTATATGCACTGGGTGGACGAAT 795
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATCGAATGATACGAACGGTCAATGATACCATGTGGCAGCTTCGTTATTAGGAATGCG 855
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 856 GCAACGAATAAAGAAATGGCACCTCAAGCGAATCTGGTTTTCATCATCATGATGATGC 915
Qy 101 AsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGCTTGGAGGCTTGCCTTCCCAATCTGCACAACTTATTCACCAAGCATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 140

Db 976 GCAGTGGCCAGATTTCATACAACTCCTGGGGGAGCGGTGAATGGGGCTACACGACA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnMetAlaValLeuPheAla 160
Db 1036 GATTCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
QY 161 GlyAsnGluGlyProAsnGlyGlyThrLeuSerAlaProGlyThrAlaLysAsnAla 180
Db 1096 GGGAAATGAAGCGGACCGGATACATCAGTGCACCTGGTACCGGTAAACCGCCATA 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnLeu 200
Db 1156 ACAGTGGGGCAACCGCAACCTCGCTCCAAAGCTTCGTTCCATGATGATGATGATGAT 1215
QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgLysProAspVal 220
Db 1216 CAGCTTGACAGTTCCTCCCGTGGCCGACAAAGATGGGCGAATCAAGCTGATGTC 1275
QY 221 MetAlaProGlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTTGCAACCGATTCCTCCTTC 1335
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrPro 260
Db 1336 TGGCGAATCATGACAGCAATATGCTATATGGGTGGACGTCCTATGGCAACCGAT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheLeuLysAsnArgLysThrPro 280
Db 1396 GTTGGGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAAATAGAGGAATCACTCT 1455
QY 281 LysProSerLeuLysAlaAlaLeuLeuAlaGlyAlaThrAspLeuGlyLeuGlyTyr 300
Db 1456 AAGCTTCCCTATTGAAAGCAGCTTTGATTGCAAGGTGCTGCTGATTTGGATTGGGTAT 1515
QY 301 ProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 1516 CCGAACGAAACCPAGGATGGGCGGAGTGCACCTCGATAAATCGTTGACGTTGCTAT 1575
QY 321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
Db 1576 GTGAACGAATCCAGTGCCTTATCACTAGCAAGCAAGCAATATACCTTTACTGCAACG 1635
QY 341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
Db 1636 CGGGGCAAGCCATTGAAATCTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1695
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
Db 1696 TCTGTAACTCCTGTCATGATTTGGATTGGTATTTGATACAGCACCAACCAAGATAT 1755
QY 381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTGGGGAATGCTTCTCAGCAACCAATTGACATTAACCTGGATGGCCCAATACGATGAA 1815
QY 401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTTATTAATTCGCCCAAGTGAACATATACCATGAGTGGTCAAGCATATAT 1875
QY 421 ValProGlnGlyProGlnAlaPheSerLeuAlaLeuValAsn 434
Db 1876 GTGGCGGTGGACCAACAACTTCTGTTGGCAATTGTGAAC 1917

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christanson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

ADDRESS: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agtis, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,81e-210 Length: 3003
Score: 2116.50 Matches: 400
Percent Similarity: 97.00% Conservative: 21
Best Local Similarity: 92.17% Mismatches: 12
Query Match: 94.07% Indels: 1
DB: 2 Gaps: 1

US-09-985-689A-6 (1-434) x US-08-873-479-41 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCCTGGCATTGTGAAGCAGACGTCGCAAAATAAATTTGGCTTAT 1529
QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAAGACAGATTGTAGCAGTTGCTGATACTGGGCTTGATACGGAAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGCATGAAGCATTCGCGGTAAGATTACCCGCACTATATGCACTGGGCGAAGCAAT 1649
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAACCCATGTTGCTGGATCTGTGTAGGAATGCT 1709
QY 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 1710 ACA--AATAAGGGATGGCCCGCAAGCCAACTCTAGCTTTTCAATCTATTATGATAGT 1766
QY 101 AsnGlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGGCTGGGAGGACTACCTCTAATCTACAAACATTTATTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValAlaTyrThr 140
Db 1827 GTTGAGCGAGATTTCACGATTCATGGGGGGTCCAGTAAGCGTGCCTATACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATTTCTTTTCGCGCC 1946

161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaIleValAla 180
1947 GGAATGAGGACCGAGTACGATCAATCACTGACCCAGGACGACCAAAATCGATT 2006
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 200
2007 ACAGTTGGGGCAACCGAAACCTACGTCGAGCTTCGATCTTATCGGATAATATTAAAC 2066
201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyValGlyIleLysProAspVal 220
2067 CATGTTGCTCAATCTCTTCACGAGGCTCTATAGAGATGAGCTATTAAAGCCGACGCTC 2126
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
2127 ATGGCACCAGGACGTATATCTCTGCTAGATCATCATATTAGCTCCAGATTCTCATTC 2186
241 TrpAlaAsnHisAspSerLysThrAlaValMetGlyThrSerMetAlaThrProIle 260
2187 TGGCAACCATGATAGTAATAGCTACATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 2246
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
2247 GTAGCAGGTAATGTTCCACATTAAGGGAGCATTTGTGAAATAGAGGGTAACTCT 2306
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr 300
2307 AAGCCTTCCTTTTAAAGCTGCTTTAATGCGAGTGTGCGGATGTTGAGCTTGGCTTT 2366
301 ProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
2367 CCAATGGTAACCAAGATGGGAGAGTAACGTTAGATAAATCCCTAATGTCGCAATT 2426
321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
2427 GTCAATGAACAGGAGCCCTTTATCAACAGTCAAAAGCAACATATTCTGTTTACGCTCAA 2486
341 SerGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrSerAla 360
2487 GCTGGTAACCCCTTAAATATCACTGTTTGTGACAGTGCACCAAGTAGACGCGCA 2546
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
2547 TCATACTTCTAGTAATGATTTAGACTTAGTAATCACTGCACCAAAATGGAATCAATAC 2606
381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGlu 400
2607 GTCGGAATGACTTTACAGCCCGTATGATAACATTTGGATGGCAGAAACACGTTGGA 2666
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
2667 AATGTTTATCAATGCTCTCAAGCGGAACGTATACAGTCGAAGTCGAGGCTTACAAT 2726
421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
2727 GTACCAAGTAGTCGCAACCTTTCTTCTAGCGATTGTACAT 2768

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/088948188
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:
Pred. No.: 4,81e-37 Length: 1977
Score: 450.50 Matches: 138
Percent Similarity: 47.05% Conservatives: 69
Best Local Similarity: 31.36% Mismatches: 170
Query Match: 20.02% Indels: 63
DB: 3 Gaps: 15

US-09-985-689A-6 (1-434) x US-08-894-818B-2 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyValValAla 27
Db 433 ATAGGGCCGATACCGTCTGGAATCTCTCGGTACGACGAGCGGTGTGGTTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db 493 ATCTGCGATACGGGTATAGACGCAAC-----CACCCCGATCTCTGAAG 534
QY 48 GlyLysIleThrAlaIleTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GCGAAGGTCTAGCTGTGTACGACCGCTCAACGCGAGTGCAGCCCTACGATGACGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLys----- 84
Db 595 GGACACGAAACCCACGTTGCGGTATCGTTCGCGAACCAGCGCAGCTTAATCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
Db 655 ATAGCGGTGCGCCCGCGCGAAGCTCGTGGCGCTCAAGGTTCTCGGTGCGAGCGTTG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAACGCTCTCCACCATCATCGCGGTGTTGACTGGGTGCTGCCAGAACAGGACAGTAC 774
QY 122 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAsp 141
Db 775 GGGATAAGGGTCAATCAACCTCTCTCGGTCTCTCCAGAGCTCCGACGGAACCGACTCC 834

US-09-985-689A-6 (1-434) x US-09-445-472-11 (1-1977)

Alignment Scores:

Pred. No.:	Length:
Score:	4,81e-37
Percent Similarity:	450.50
Best Local Similarity:	47.05%
Query Match:	31.36%
DB:	Indels: 170
	Gaps: 63

US-09-985-689A-6 (1-434) x US-09-445-472-11 (1-1977)

142 SerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGly 161
835 CTCAAGTCAGCGCGTCAACAAAGCGCTGGAGCGCGGTATAGTAGTCTGCTGCGCGCGGC 894
162 AsnGluGlyProAsnGlyGlyThrLeuSerAlaProGlyThrAlaLysAsnAlaLeuThr 181
895 AACAGCGCGCGGACCACTACACCGTGGCTACCGCGCGCGCGCGCGCGCGCGCGCGCG 954
182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnLeuAsnHis 201
955 GTCCGTGCA-----GTTGACAGCAACGACAAC 981
202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyValArgLeuValMet 221
982 ATGCCAGCTTCTCCAGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1041
222 AlaProGlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerPheTyr 241
1042 GCCCGCGCGCTGACATCATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1092
242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProLeuVal 261
1093 ACCCGGATAACGACTACTACCAAGCGCTCTGGAAACCGAGCATGCGCAACCGCGCGCT 1152
262 AlaGlyAsnValAlaGlnLeuArgGluHisPheLeuLysAsnArgGlyLeuThrPro--- 280
1153 TCGGCGCTGGCGCGCTCATCTCTCCAG-----GCCACCGAGCTGGACCGCGGAC 1203
281 -----LysProSerLeuLeuLeuAlaAlaLeuLeuAlaLeuAlaLeuAlaLeu 297
1204 AAGGTGAAGACCGCGCTCATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1263
298 LeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu--- 316
1264 ATCGCCTAC-----GTTGCGGTAGGTGAACTCTACAGGCGCATCAAG 1308
317 -----AsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr 334
1309 TACGACGACTACGCGCAAGCTACCTTACCGCGCTCGTGGCGCGCAAGGGAAGCGCGCAC 1368
335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysLeuSerLeuValTyrSerAspAla 354
1369 CACACTTCGACGCTCAGCG 1422
355 ProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValLeuThrAla 374
1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTACGAC 1458
375 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAsp 394
1459 CCCAACGGGAACGAG-----GTTGACTACTCTACACCGCGCTACTAC----- 1500
395 GlyArgAsnValGluAsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrVal 414
1501 -----GGCTTCGAGAGGTGGCTACTACACCGCGCGCGCGCGCGCGCGCGCGCG 1551
415 GluValGlnAlaTyrAsnValProGlnGlyProGlnAlaPheSerLeuAlaLeuValAsn 434
1552 AAGGTGCTGACGCTAC-----AAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1602

RESULT 6
US-09-445-472-11
Sequence 11, Application US/09445472
Patent No. 638726
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikuroshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA=6

US-09-985-689A-6 (1-434) x US-09-445-472-11 (1-1977)

Alignment Scores:

Pred. No.:	Length:
Score:	4,81e-37
Percent Similarity:	450.50
Best Local Similarity:	47.05%
Query Match:	31.36%
DB:	Indels: 170
	Gaps: 63

US-09-985-689A-6 (1-434) x US-09-445-472-11 (1-1977)

8 VallysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnValValAla 27
433 ATAGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGAGCGGTGTGTGTGTC 492
28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
493 ATCTGTCATACCGGTATAGACGGAAC-----CACCCGATCTGAAG 534
48 GlyLysIleThrAlaIleTyr-----AlaLeuGlyArgThrAsnAlaAsnAspProAsn 66
535 GGCAAGGTCATAGGTGTCGACCGCTCAACGGCAGGTTCGCGGACCGGCGGTAACTCCAGTAC 594
67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLys----- 84
595 GGACACGCGACCAACCGCTGCGGGTATCGTTCGCGGACCGGCGGTAACTCCAGTAC 654
85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
655 ATAGGCGTCCCGCGCGCGGAGCTCGCGGTCAAGGTTCTCGGTGCGCGCGGTTCG 714
102 GlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
715 GGAAGCGTCCACCATCATCGCGGTGTGACGCTGCGTCCGACGACGACGACGACGACGAC 774
122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAsp 141
775 GGGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGACGACGACGACGAC 834
142 SerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGly 161
835 CTCAAGTCAGCGCGTCAACAAAGCGCTGGAGCGCGGTATAGTAGTCTGCGTCCGCGCG 894
162 AsnGluGlyProAsnGlyGlyThrLeuSerAlaProGlyThrAlaLysAsnAlaLeuThr 181
895 AACAGCGCGCGGACCACTACACCGTGGCTACCGCGCGCGCGCGCGCGCGCGCGCGCG 954
182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnLeuAsnHis 201
955 GTCCGTGCA-----GTTGACAGCAACGACAAC 981
202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyValArgLeuValMet 221
982 ATGCCAGCTTCTCCAGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1041
222 AlaProGlyThrPheLeuSerAlaArgSerSerLeuAlaProAspSerPheTyr 241
1042 GCCCGCGCGCTGACATCATAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1092
242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProLeuVal 261

1093	DB	ACCCGATAAACGACTACTACACCAAGGCGCTCGGAACCAAGCATGGCCACCCCGCAGCTT	1152
262	QY	AlaGlyAsnValAlaGlnLeuArgGluHisPheIleIysAsnArgGlyIleThrPro---	280
1153	DB	TCGGGCGTTGGCGCGCTCATCTCCAG-----GCCACCCGAGCTGGACCCCGCAG	1203
281	QY	-----LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGly---	297
1204	DB	AAGGTGAAGACCCGCTCATCGAGACCGCGACATATCGCCCCCAAGGATAGCGGAC	1263
298	QY	LeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu---	316
1264	DB	ATCGCCTAC-----GGTGGGFTAGGGTGAAGCTCTACAGGCCATCAAG	1308
317	QY	-----AsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr	334
1309	DB	TACGACGACTACGCCAAGCTCACCTTACCGGCTCGTCCGCGACAAAGGACGGCCACC	1368
335	QY	TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValThrSerAspAla	354
1369	DB	CACACTTCGACGTACGCGGCGCACCTCGTACCGCCACCCCTCTACTGGAC-----	1422
355	QY	ProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAla	374
1423	DB	-----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGC	1458
375	QY	ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp	394
1459	DB	CCCAACGGGAACGAG--GTTGACTCTCTACACCGCCTACTAC-----	1500
395	QY	GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrVal	414
1501	DB	-----GGCTTCGAAGAGTCGGTACTACACCCGACCGCGGAACCTCGACGGTC	1551
415	QY	GluValGlnAlaTyrAsnValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn	434
1552	DB	AAGTCTCTCAGCTAC-----AAGGGCGGCGCACTACACAGGTCGACGTCGTCAGC	1602

US-09-985-689A-6 (1-434) x US-09-445-472-2 (1-1236)

Qy	12	ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl	24
Db	24	GTCTGCAGCTCAAGTTATTCGGCAACTTACGTTTGGAACTTGGGATATGATGGTTCTCGAAT	83
Qy	24	nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl	44
Db	84	CACATATAGGAATAATTGACACTGGAAATTGAC-----GCITTCATCC	125
Qy	44	uAlaPheArgGlyLysIleThrAlaIleTyrAlaIleuGlyArgThrAsnAsnAlaAsn--	63
Db	126	AGATCTCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATCG	170
Qy	64	-----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGl	78
Db	171	TAGGAGTTATCCATACGATGACCATGGACATGGAACCTCATGTAGCTTCAATAGCAGCTCG	230
Qy	78	yAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh	94
Db	231	TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCAGGAGCTTAAGCTCGCGG	290
Qy	94	eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValse	112
Db	291	AATTAAAGTTCTAGGTGCGGATGTTCTGGAAGCATATCTACTATAATTAAGGAGATTGA	350
Qy	112	rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl	132
Db	351	GTGGCGCGTTGTATPAAACAAAGATAAGTACGGAATTAAGGTCATTAACTTCTTCTTCTG	410
Qy	132	aProValAsnGlyValAlaTyrThrThrAspSerArgAsnValaAspAspTyrValArgLysAs	152
Db	411	AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGCAGCGTGGGATGC	470
Qy	152	nAspMetAlaValLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyIleThrIleSerAl	172
Db	471	TGGATTAGTTGTTGTGTGCTGCCCTGGAACAGTGCACCTTCAAGCATACATCGGTT	530
Qy	172	aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh	192
Db	531	TCCAGCAGCTGCAAGCAAAGTTATTACAGTTGGAGCC-----567	567
Qy	192	eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly	212
Db	568	-----GTTGCAAGTATGATGTTATACAAAGCTTCTCAAGCAGAGGGCCAACTGC	617
Qy	212	sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgse	232
Db	618	AGACGGCAGCGCTTAAGCCTGAGGTGTGCTCAGGAAACTGGATAATTGCTGCCAGAGC	677
Qy	232	rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGl	252
Db	678	AAAGT-----GGAACTAGCATGGCTCAACCAATTAATGATATTACACAGCAGCTCC	728
Qy	252	yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh	272
Db	729	TGGGACATCAATGGCAACTCCTCAGCTAGCTGTTGTTATTCAGCCCTCTTGTCCAA	783
Qy	272	eIleLysAsnArgGlyIleThrPro-----LysProSerLeuLeuLysAlaIle	289
Db	784	----GCACACCCGAGCTGGACTCAGACAAGATAAAACAGCCCTCATAGAACTGCTGA	839
Qy	289	uIleAlaGlyAlaThrAspIleGly---LeuGlyTyrTyrProSerGlyAsnGlnGlyTrpGl	308
Db	840	TATCGTAAAGCCAGATGAAATAGCCGATATAGCTAC-----GCTGCAGG	884
Qy	308	yArgValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe	325
Db	885	TAGGTTTAATGCATACAAAGCTATAAACTAGTAACTATGCAAGCTAGTTCTACTCG	944
Qy	325	rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe	345
Db	945	ATATGTTGCCAACAAAGGACGCCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTCGT	1004

QY 345 uylsileSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATCTACTGGACAAATGCCAAT----- 1035
QY 365 llaenAspLeuAspLeuValleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACTCTACATCCCAATGGAAACCCAG---GTTGACTACTCTTA 1091
QY 385 eThralaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheileAs 405
Db 1092 CACCGGCTACTAT-----GGATTCGAAAGGTTGGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1128 CAACCCAACTGATGGACATGGCAATTAAGTTGTAAGCTACAGC-----GGAAG 1178
QY 425 oGlnAlaPheSerLeuAlaileValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 8

US-08-894-818B-4
; Sequence 4, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.

US-08-894-818B-4

Alignment Scores:
Pred. NO.: 1.28e-32 Length: 1566
Score: 406.50 Matches: 136
Percent Similarity: 44.22% Conservative: 63
Best Local Similarity: 30.22% Mismatches: 169
Query Match: 18.07% Indels: 82
DB: 3 Gaps: 17
US-09-985-689A-6 (1-434) x US-08-894-818B-4 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCAGCTCAAGTTATGGCAACTTACGTTGGAACTTGGGATATGATGTTCTGGAAT 83
QY 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGGAAATTGAC-----CCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysileThrAlaileTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAGGANAAGTA-----ATTGGGTGGGTAGATTGTCATGG 170
QY 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACATGACATGACATGACATGACATGACATGACATGACATGACATG 230
QY 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGACGACAGTAAGTATGCAAGTACCAAGGGAATGGCTCCAGGAGCTAAGCTGGCGG 290
QY 94 eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSe 112
Db 291 AATTAAGGTTCTAGGTGCCGATGTTCTGGAAGCATATCTACTATAATTAAAGGAGGTGA 350
QY 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyValaArgileHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCCCTTGTATCAACAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTTGTTTC 410
QY 132 aProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValaArgLysAs 152
Db 411 AAGCCAGAGCTCAGATGTTACTGACGCTCTAAGTCAGGCTGTTAATCAGCGTGGGATGC 470
QY 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrileSerAl 172
Db 471 TGGATTAGTTGTTGGTTGGCTGGAAACAGTGGACCTAACCAAGTATACAATCGGTTTC 530
QY 172 aProGlyThrAlaLysAsnAlaileThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 531 TCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGAGCC----- 567
QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 568 -----GTTGACAAGTATGATGTTATTAACAGCTTCTCAAGCAGAGGGCCAACTGC 617
QY 212 sAspGlyArgileLysProAspValMetAlaProGlyThrPheileLeuSerAlaArgSe 232
Db 618 AGACGGCAGGCTTAAGCCTGAGGTTGTTGCTCCAGGAAACTGGATAATTGTCGCCAGAGC 677
QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetG1 252
Db 678 AAGT-----GGAACATGATGCTCAACCAATTAATGACTATTACAGCAGCTCC 728
QY 252 yGlyThrSerMetAlaThrProileValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 729 TGGGACATCAATGGCAACTCTTCACGTAGTGTGTTGCGCCCTCTTGCTCCAA----- 783
QY 272 eileLysAsnArgGlyileThrPro-----LysProSerLeuLeuLysAlaAla 289
Db 784 ----GCACCCCGAGCTGACATCCAGACAAAGTAAACAGCCCTCATCAAGCTCTGA 839
QY 289 uileAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTrpG1 308


```
Db 840 TATCGTAAAGCCAGATGAATAGCCGATATAGCCTAC-----GGTGCAGG 884
Qy 308 YARGValThrLeuAspLysSerLeu-----AsnValAlaPheValaenGluThrSe 325
Db 885 TAGGTTAATGCATACAGGCTATAACTACGATACATCAAGCTAGTGTCTACTGG 944
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrsrPheThrAlaGlnSerGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGAGCCAACTCACCAGTTGCTTATTAGCGGAGCTTCGTTGCT 1004
Qy 345 uLysleSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATATATGGCAATGCCAAT----- 1035
Qy 365 laenAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACAG--GTTGACTACTCTTA 1091
Qy 385 eThrAlaProTyrrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCCTACTAT-----GGATTGCAAAAGGTTGCTTATTA 1127
Qy 405 rAlaProGlnSerGlyThrTyrrValGluValGlnAlaTyrrAsnValProGlnGlyPr 425
Db 1128 CAAACCACTGATGAAATGACGACAAATTAAGTTGTAAGCTACAGC-----GGNAG 1178
Qy 425 oglnAlaPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGCTGTAAGT 1206
```

RESULT 9

```
US-08-894-818B-34
Sequence 34, Application US/088994818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
```

```
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34
Alignment Scores:
Pred. No.: 1,846-32 Length: 1962
Score: 406.50 Matches: 136
Percent Similarity: 44.22% Conservative: 63
Best Local Similarity: 30.22% Mismatches: 169
Query Match: 18.07% Indels: 82
DB: 3 Gaps: 17
US-09-985-689A-6 (1-434) x US-08-894-818B-34 (1-1962)
Qy 12 ValAlaGlnSerSerTyrrGly-LeuTyrr-----GlyGlnGlyCl 24
Db 420 GTCTGCAGCTCAAGTTATGGCACTTACGTTTGGAACTTGGGATATGATGGTCTCGAAT 479
Qy 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethISG1 44
Db 480 CACAATAGGATATATTGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyrrAlaLeuGlyArgThrAsnAsnAlaSer-- 63
Db 522 AGATCTCCAAAGGAAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
Qy 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGATTATCCATACGATGACCATCGACATGGAATCACTCATGTAGTTCATAGCAGCTGG 626
Qy 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGGAATGGCTCCAGGAGCTAAGCTGCGG 686
Qy 94 eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSe 112
Db 687 AATTAGGTTCTAGTGCCGATGGTCTTGGAGCATATCTACTATATTAATTAAAGGAGTTGA 746
Qy 112 rThrLeuPheSerGlnAlaTyrrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCGCGTTGATAACAAAGATAAGTACGGAATTAAGTCAATTAATCTTCTTCTGTTTC 806
Qy 132 aProValAsnGlyAlaTyrrThrThrAspSerArgAsnValAspAspTyrrValArgLysAs 152
Db 807 AAGCCAGAGCTCAGATGCTGACTCAGCGCTTAAGTCAGGCTGTTAATGCGAGCGGGATGC 866
Qy 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAl 172
Db 867 TGGATTAGTTGTTGTTGCTGCGCTGGAAACAGTGGACCTAACCAAGTATACATCGGTTTC 926
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 TCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGAGCC----- 963
Qy 192 eGlySerTyrrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 964 -----GTTGACAAAGTATGATGTATACAGCTTCTCAAGCAGAGGCGCAACTGC 1013
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1014 AGACGGCAGGCTTAAGCCTGAGTTGTTGCTCCAGGAAACTGGAATATGCTGCCAGAGC 1073
Qy 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrrAlaTyrrMetG1 252
Db 1074 AAGT-----GGAATAGCATGGGTCAACCAATTAATGACTATATACAGCAGCTCC 1124
Qy 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
```

Db 1125 TGGACATCAATGGCACTCTCACGTAGCTGGTATTGCGAGCCCTTGTCTCCAA----- 1179
Qy 272 eilelyAsnArgGlyIleThrPro-----LysProSerLeuLeuLysAlaAlaLe 289
Db 1180 ----GCACACCGAGCTGACTCCACAAAGTAAACAGCCCTCATAGAACTGCTGA 1235
Qy 289 uileAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTyrPgl 308
Db 1236 TATCGTAAAGCCAGATGAATAGCCGATATAGCTTAC-----GGTGCAGG 1280
Qy 308 yArgValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGGTTAATGCATACAGCTATAAAGTACAGTAACTATGCAAGCTAGTCTTCACTGG 1340
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGAGCCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1400
Qy 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 AACTGCCATATTACTGGGACCAATGCCAA----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1487
Qy 385 eThrAlaProTyrAspAsnAsnTyrPaspGlyArgAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCTACTAT-----GGATTGGAAGGTTGGTTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1524 CAACCCAACTGATGGACATGGACAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
Qy 425 oGlnAlaPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGCTGTAAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15 Application US/09445472
; Patent No. 6358726

; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/09/445,472

; CURRENT FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997

; PRIOR FILING DATE: 1997-06-10

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 15

; LENGTH: 1962

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic

US-09-445-472-15

Alignment Scores:

Pred. No.: 1,84e-32 Length: 1962

Score: 406.50 Matches: 136

Percent Similarity: 44.22% Conservative: 63

Best Local Similarity: 30.22% Mismatches: 169

Query Match: 18.07% Indels: 82

DB: 4 Gaps: 17

US-09-985-689A-6 (1-434) X US-09-445-472-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTCGAAT 479
Qy 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethHisGl 44
Db 480 CACAATAGGAATTAATGACACACTGGAAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAAAGTA-----ATTGGTGGTAGATTTTCTCAATGG 566
Qy 64 -----AspProasnglyHisGlyThrHisValAlaAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATPACCATGACATGCAATGGAACATCATGTAGCTTCAATACACACTGG 626
Qy 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGCGGG 686
Qy 94 eGlnSerValMet-----AspSerAsnGlyGlyLeuGlyGlyLeuProSerAsnValSe 112
Db 687 AATTAAAGGTTCTAGGTCCCGATGGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
Qy 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
Db 747 GTGGCGCGTTGATACAAAGTAAGTACGGAATTAAGTTCATTAATCTTCTCTGTTTC 806
Qy 132 aProValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAs 152
Db 807 AAGCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGCTGTTTAATGACAGCGTGGGATGC 866
Qy 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAl 172
Db 867 TGGATTAGTTGTTGTTGCTGCGCTGGAACAGTGGACCTTACCAAGTATACATCGTTTC 926
Qy 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 TCCAGCAGCTGCAAGCAAGTTATTACAGTTGGAGCC----- 963
Qy 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 964 -----GTTGACAGTATGATGTTATAACAGCTTCTCAAGCAGAGGGCCAACTGC 1013
Qy 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1014 AGACGCGAGGCTTAAGCCTGAGTTGTTGCTCCAGGAACTGGATAATTGTCGCAGAGC 1073
Qy 232 rSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGl 252
Db 1074 AAGT-----GGAAGTACATGGGTCAACCAANTTAATGACTATTACACAGCAGCTCC 1124
Qy 252 yGlyThrSerMetAlaThrProIleValAlaAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 1125 TGGGACATCAATGGCAACTCTCTCACGTAGCTGGTATTGTCAGCCCTCTTGCTCCAA----- 1179
Qy 272 eilelyAsnArgGlyIleThrPro-----LysProSerLeuLysAlaAlaLe 289
Db 1180 ----GCACACCGAGCTGAGCTCCAGACAAAGTAAACACAGCCCTCATAGAACTGCTGA 1235
Qy 289 uileAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTyrPgl 308
Db 1236 TATCGTAAAGCCAGATGAATAGCCGATATAGCTTAC-----GGTGCAGG 1280
Qy 308 yArgValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGGTTAATGCATACAGGCTATAAAGTACAGTAACTATGCAAGCTAGTCTTCACTGG 1340
Qy 325 rSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGAGCCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1400

QY 345 ulysileSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATCTGGGCAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrIysTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTCGATCCCAATGGAACACAG--GTTGACTACTCTTA 1487
QY 385 eThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCCTACTAT-----GGAATTCGAAAGGTTGGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGlnGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGGAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
QY 425 oglnAlaPheSerLeuAlaIleValAsn 434
Db 1575 TCGAACTATCAAGTAGATGTTGTAAGT 1602

RESULT 11

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Broadway and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA: JP 323285/1995
; APPLICATION NUMBER: 12-DEC-1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-894-818B-6

Alignment Scores: 2.37e-32 Length: 1977
Pred. No.: 405.50 Matches: 134
Score: 45.11% Conservative: 69
Percent Similarity: 29.78% Mismatches: 165
Best Local Similarity: 18.02% Indels: 82
Query Match: 3 Gaps: 17
US-09-985-689A-6 (1-434) x US-08-894-818B-6 (1-1977)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGAACCTTGGGATATGATGTTCTGAAT 479
QY 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAGAAGTA-----ATTGGTGGGTAGATTGTTCAATGG 566
QY 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGACATGGAAGTCACTAGCTTCAATAGCAGCTGG 626
QY 78 YAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAGTAATGCGAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerValMet-----AspSerAsnGlyLeuGlyGlyLeuProSerAsnValse 112
Db 687 AATTAAGGTTCTAGTGCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGCTGA 746
QY 112 rThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCGCTTGATAACAAAGATAGTACGGAATAAGAGTCACTAAATCTTCTTCTGCTTC 806
QY 132 aProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAs 152
Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGGCGGTCAACAACGCTGGGACGC 866
QY 152 nAspMetAlaValLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAl 172
Db 867 CGGTATAGTAGTCTGCGTCGCGCGCAACACGCGGCGCAACCTACACCGCTCGGCTC 926
QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 927 ACCGCGCGCGGAGCAGGAGTCAACCGTCCGTGCA----- 963
QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 964 -----GTTGACAGCAACACGACACATCGGCAGCTTCTCCAGCAGGCGGACCGCCG 1013
QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgse 232
Db 1014 GGACGGAAGGCTCAGCGGAGTGTGTCGCCCGCGGCTTGACATCATAGCCCCGCGCGC 1073
QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrTyrAlaTyrMetG1 252
Db 1074 CAGC-----GGAACCCAGCATGGGCACCCCGATAAACGACTACTACACCAAGGCGCTC 1124
QY 252 yGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPh 272
Db 1125 TGGAAACACGATGGCCACCCCGCAGCTTTCGGGCGTTGGCGGCTCATCTCCAG----- 1179
QY 272 eIleLysAsnArgGlyIleThrPro-----LysProSerLeuLysAlaAla 289
Db 1180 -----GCCACCCGAGCTGGACCCCGGACAAAGGTGAAGACCGCTCATCGAGACCGCGCA 1235
QY 289 uIleAlaGlyAlaThrAspIleGly---LeuGlyTyrProSerGlyAsnGlnGlyTrpG1 308
Db 1236 CATAGTCGCCCCCAAGGAGATAGCGGACATCGGCTAC-----GGTGGCGGG 1280

```
QY 308 YATGValThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSe 325
Db 1281 TAGGTCGAAGCTTACAGGCCATCAAGTACGACGACTACGCAAGCTCACCTTACCGG 1340
QY 325 rSerLeuSerThrAsnGlnLysAlaThrTySerPheThrAlaGlnSerGlyLysProLe 345
Db 1341 CTCGTCGCGCAGAACAGGAGCGCCACCCACACCTTCGACGTCAGCGCGGCCACCTTGT 1400
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuVa 365
Db 1401 GACCGCCACCTTCTACTGGAC-----ACGGGCTC 1430
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPh 385
Db 1431 GAGCGACATCGACTCTACTCTACGACCCCAACGGAACGAG---GTTGACTACTCTCTA 1487
QY 385 eThrAlaProTyAspAsnAsnTrpAspGlyAAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTACTATC-----GGCTTCGAAGTCTGGTACTA 1523
QY 405 nAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnValProGlnGlyPr 425
Db 1524 CAACCCGACCGCGGAACCTGAGCGTCAAGGTCTGTCAGCTAC-----AAGGCGC 1574
QY 425 cGlnAlaPheSerLeuAlaIleValAsn 434
Db 1575 GCGGAACCTACGAGTGCAGCTGCTCAGC 1602

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
```

```
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
US-09-000-016-3

Alignment Scores:
Pred. No.: 1.74e-27 Length: 2539
Score: 360.50 Matches: 146
Percent Similarity: 43.38% Conservative: 57
Best Local Similarity: 31.20% Mismatches: 175
Query Match: 16.02% Indels: 90
DB: 3 Gaps: 22

US-09-985-689A-6 (1-434) x US-09-000-016-3 (1-2539)
QY 2 AspValAlaAaGGGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGly 21
Db 953 GACACGTCCTCGCGGCAAGATCGGCGCCGCCCAAGCGTGTCTCCGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTCGACACCGGTCTCGACACGAGC----- 1054
QY 42 MethHisGluAlaPheArgGlyIleThrAlaIleTyAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGACCTGAAGGCGCGGTGACCGCTCCCAAGAACTTCACCGCGCCCGCC 1111
QY 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCCGCGACAAAGTGGGCGCACCGCACCGCTCGCTCGATCGCGGGGCGACGGCGCC 1171
QY 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCAAAGGCAAGTACAAAGGGGTGCGACCCCGCGCGGATCTCTCAACGGCAAGTCT 1231
QY 98 MetAspSerAsnGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGCT-----TTCCGGACACACTCCGCGCATCTCCCGGCGATGAGTGG 1285
QY 118 AlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 GCGGCGCGCGAGCGCGAGCTGTCCACCATGAGCTCTGGGC-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspTyValArgLys-----AsnAspMetAla 155
Db 1334 GACACACCGGAGACCGCCGCTGAGCGCGGCTCGACAAGCTTCCCGCGAGNAGGGC 1393
QY 156 ValLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTTGTTCGCCATCGCGCGCGGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGACGCGCGCTCTACCGTCCGCGCC----- 1480
QY 194 SerTyAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAAAGGACGACTCCCGACTTCTCTCCACCGCGCCCGCTCGGC 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGCCATCAAGCGGACGTCACCGTCCCGCGGTGGACATCACCGCGCTCCGCGG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGly 252
Db 1595 GAGGGCAACGACATCGCGCAGGAGTCTGAGGGACCGCGGTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
```

Db 1655 GGCACGTCGATGGCGACCCCGCAGTCGCGGGCGCGGCCCTCTCTGAAGCAGCAG--- 1711
Qy 273 IleIysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaIleLeu----- 289
Db 1712 -----CACCCCGACTGGACCTCCGCCCACTGAAGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACCGCTCCACCAAG---GGCGCAAGTACACCCGTTCCAGCAGGTTCCGGCCCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerLeuSer--- 328
Db 1805 ATCCAGGCCGACAAAGGCTCCAGCAGCCGTGATCGCCAGCCCGTCTCGGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GGGTCACAGTGGCGGCACACGACGACGCGGTACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTGGCACCCAGGACGTCACGCTGACGTCGACCGCCACCGACCCCAAGGCG 1984
Qy 354 -----AlaProLaser-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGCGCCCGCGCGCTTCTTACGTCGGCGCCACACACG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGCTCCCGCGCGCGGCGAGC---GCCTCCGTCGACATGACCGCC----- 2071
Qy 390 AspAsnAsnTrpAspGlyArg-----AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCCGGCTCGCGCACCGGTGGACGCGCGTACTCGCGTACCTGGTTCGCGC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGCGGCGAGACGGTCCSCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCGGGCACATCGGCC 2210

RESULT 13
US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akita ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng

REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3

Alignment Scores:
Pred. No.: 1.74e-27 Length: 2539
Score: 360.50 Matches: 146
Percent Similarity: 43.38% Conservative: 57
Best Local Similarity: 31.20% Mismatches: 175
Query Match: 16.02% Indels: 90
DB: 4 Gaps: 22

US-09-985-689A-6 (1-434) x US-09-514-340-3 (1-2539)

Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnInsersertYrGlyLeuTyrGly 21
Db 953 GACAGCTCCGTCGGGAGATCGGCGCCCAAGCGGTGTCGCC---GGCTACGACGGC 1009
Qy 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGTGAAGATCGCGCTCTGACACCGCGTCTGCACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGCGCGGTACCGCTCCCAAGACTTCACCGCGCGCGCGC 1111
Qy 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCGCGCGCAAGTGGGCGCACCGCACCGCCTCGATCGCGCGCGCGCGCGCGCGCGC 1171
Qy 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCAAAGGCAAGTACAGGCGCTGCACCCCGCGCGCGCTCTCAACGCGCAAGTC 1231
Qy 98 MetAspSerAsnGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGCT-----TTCGGCGACGACTCCCGCATCTCCCGCGCATGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 GCGCGCGCGCGCGCGCGCGCTGTCACCATGAGCTGGGC-----GGCATG 1333
Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetAla 155
Db 1334 GACACCGGAGACCGCCCGCTGGAGGCGCGCTGCACAGCTGTCGCGCGAGAGGCGC 1393
Qy 156 ValLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTTTCGC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GCGAGCGCGCGCGCGCGCGCTCCTCCTCGTGGCGCC----- 1480

194	SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys	212
1481	-----GTCCAGCAAGAAAGTCCGCCGACTTCTCTCCACGGGCCCGCGCTCGGC	1534
213	AspGlyArgIleValProAspValMetalProGlyThrPheIleLeuSerAlaArgSer	232
1535	GACGGCGCATCAAGCGGACGTACCGTCCCGCGTGGACATACAGCGCGCTTCGGCG	1594
233	SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly	252
1595	GAGGCAACGACATCGCGCAGGAGTCCGTGAGGACCGCGCGGTACATGACCATCTCC	1654
253	GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe	272
1655	GCACCGTCATCGGACCGCACGTCGCGGGCGCGCGCCCTCTCGAAGCAGCAG---	1711
273	IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----	289
1712	-----CACCCCGACTGGACCTCCGCGCACTGAAGGCGCGG	1747
290	IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArg	309
1748	CTCACCGGTCCACCAAG---GGCGGCAAGTACACCCGCTTCGAGCAGGTTTCGGGCCG	1804
310	ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeuSer---	328
1805	ATCCAGGCGCACAAAGCGCTCCAGCAGACCGGTATGCCGACCCGGTCTCGGTAGCTTC	1864
329	-----ThrAsnGlnLysAlaThrTyrSer	336
1865	GGCGTCCAGCAGTGGCGCACACGAGCAGCGCGTCAACCAAGCAGCTGACTACCGC	1924
337	PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp-----	353
1925	AACCTCGGACCCAGGACGTACGCTGAAGTACGTCGCGCCGCCACCCACCCAAAGGC	1984
354	-----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp	369
1985	AAGCGGCCCGCGGGGCTTTCACGCTGGCGGCCACACG-----	2026
370	LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr	389
2027	-----GTGACCGTCCCGGGGGCGGACG---GCCTCCGTCGACATGACCGCGC---	2071
390	AspAsnAsnTrpAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln	408
2072	GACACCGCGTCCGGCGCACCGTGTGACGGCGCGTACTCGCGGTACGTGGTGCGC---	2128
409	SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn	420
2129	GGCGGGCGGACAGCGTTCGCACGCGCGCGCGGTGCAGCGCAGGTGCATGCTACGAC	2188
421	Val-ProGlnGlyProGlnAla	427
2189	GTGACCGTCCGGCACATCGGCC	2210

RESULT, T 14

RECORDED 14
 US-09-000-016-1
 ; Sequence 1, Application US/09000016
 ; Patent No. 6143541
 ; GENERAL INFORMATION:
 ; APPLICANT: AKIRA ARISAWA et al.
 ; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
 ; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
 ; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 ; STREET: 2033 K Street, N.W., #800
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 6,77e-27
Score: 355.50
Percent Similarity: 43.16%
Best Local Similarity: 30.98%
Query Match: 15.80%
DB: 3
US-09-985-688A-6 (1-434) x US-09-000-016-1 (1-2)

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000.016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1

Alignment Scores:
Pred. No.: 6 77e-27 Length: 2809
Score: 355.50 Matches: 145
Percent Similarity: 43.16% Conservative: 57
Best Local Similarity: 30.98% Mismatches: 176
Query Match: 15.80% Indels: 90
DB: Gaps: 22

US-09-985-689A-6 (1-434) x US-09-000-016-1 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGly 21
Db 953 GACACGTCCGTCGGGCAGATCGCGCCGCCCAAGCGTGTCGCCGC---GGTACGACGCG 1009
QY 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTCGACACGCGTGTGCACACGACG----- 1054
QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaIleTyAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGGCGCGGTGACCGCGTCCACAGACTTCACCGCGCGCGCCGCGC 1111
QY 62 AlaAsnAppProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCGCGCGCAAGGTGGGCGCACGCCACCGTCCGCTCGATCGCGGCGGACGCGGCGCC 1171
QY 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCAGGCAAGTACAAAGGGCGTCGCAACCGCGCGCGCATCTCAACGGCAAGGTC 1231

```

Qy 98 MetAspSerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCGGCGAGGACTCCGGCATCTCCGGCGCATGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 GCGCGCGCGAGGCGCGGCGTCAATGAGCTGGGC-----GGCATG 1333
Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgIle-----AsnAspMetAla 155
Db 1334 GACACCGGAGAGCCAGCCGCTGGAGGGCGGCTGACAGAGTGTCCCGCCGAGAGGCG 1393
Qy 156 ValLeuPhe-----AlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTTCGCATCGCGGCGCGCAAGAGGGCGCGAG-----TCGATCGTTCCGCC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGAGCCCGCCCTCACCGTCGGCGCC----- 1480
Qy 194 SerTyrAlaAspAsnIleAsnHisAlaAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAGGACAGCTCGCGACTTCTCTCCACCGCGCCCGCTCGCG 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCCATCAAGCGGACGTCACCGCTCCCGGGCTGGACATCACGCGCGCTCGCG 1594
Qy 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGCAACGACATCGCGCAGGAGTGGTGGAGGACCGCGCGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCGATGGGACCGCGACGTCGCGGGCGCGCGCTCTCTGAGCAGCAGCAG--- 1711
Qy 273 IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGACCTGGACCTCCCGCCGACTGAAGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGCTCCACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGCTTCGGGCGG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeuSer--- 328
Db 1805 ATCCAGGCGCAAGAGGCGTCCACGACCGGTGATCCCGCCGCGTCTCGGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCAGTGGCGCACACGACGACGCGGCTCACCAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGACCCAGGAGCTCACGCTGAAGCTGACGTCGACCGCCACCGCCCAAGGCG 1984
Qy 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGCGCCCGCGGGCTTCTTACGCTGGCGCCACCG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGCGGCGGCGAGC---GCCTCGCTGACATGACCGCC----- 2071
Qy 390 AsnAsnAsnTrpAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCGCGTGGCGGCGACGCTGACCGCGTACTCGCGTACGTGGTGGCGCC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGCGGCGACGCTCCGCGCGCGCGCGGCTGACGCGGAGGTGAGTCTGCTACGAC 2188
Qy 421 Val-ProGlnGlyProGlnAla 427

Db 2189 GTGACCGTCCGGCAGCATCGGCC 2210
RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akita ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE 1
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338..2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores: 6.77e-27 Length: 2809
Pred. No.: 355.50 Matches: 145
Score: 35.50
Percent Similarity: 43.16% Conservative: 57
Best Local Similarity: 30.98% Mismatches: 176
Query Match: 15.80% Indels: 90
DB: 4 Gaps: 22
US-09-985-689A-6 (1-434) x US-09-514-340-1 (1-2809)
Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21

Db 953 GACACGTCCGTGGGCGAGATCGGGCCCCCAAGCGGTGTCTCGCC---GGCTACGACGGC 1009
Qy 22 GlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGGCGTGAAGATCGCGCTCTCGACACCGGTCTCGACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGGCGGGTGCACCGCTCCAGAACTTCACCGCGCGCCCGC 1111
Qy 62 AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThr 81
Db 1112 GCCCGGACAAAGTGGCGCCACGGACCCACGTCGCTCGATCGCGCGGGGACCGGGCGC 1171
Qy 82 SerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 1172 CAGTCCCAAGGCAAGTAGACAGGGGTGCACCGCGCGCGGATCTCTCAACGCGCAAGGTC 1231
Qy 98 MetAspSerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCGGGACGACTCCGGCATCTTCGGCGGCATCGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAla 137
Db 1286 CGGCGCGGCGAGGGCGCGAGCTCGTCAACATGAGCTGGGC-----GGCATG 1333
Qy 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetAla 155
Db 1334 GACACACGGAGACCGCCCTGGAGGGCGGGTTCGACAGCTGTCCGCGGAGAGGGC 1393
Qy 156 ValLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCGTGTCGCATCGCGCGCGGCAACGAGGCGCGGAG-----TCGATCGTTCGCC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GCGACGGCGGACGCGCGCTCACCCTCGCGGC----- 1480
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAGGACAAGCTCGCGGACTTCTCTCCACCGGCGCGGCTCGGC 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCATCAAGCGGACGTCCCGCTCCGGCGTGGACATCACGGCGCGCTCGGCG 1594
Qy 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGCAACGACATCGCCGAGGAGTCTGGTGGGAGCGCGGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCGATGGCGACCGCGCACGTCGCGGGCGCGGCGCTCTCTGAAGCAGCAG--- 1711
Qy 273 IleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGACTGACCTCCGCGCACTGAAGGGCGCG 1747
Qy 290 IleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGGCTCCACCAAG---GGCGGCAAGTACACCCGCTTCGAGCAGGGTTCGGGCGCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerSerLeuSer--- 328
Db 1805 ATCCAGGCCGACAAAGGCGCTCCAGCAGACCGTGTATCGCGACCGCGGTCTCGGTGAGCTTC 1864
Qy 329 -----ThrAsnGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCAGTGGCGGCACACCGACGACGAGCGGTACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCCAGGACGTACGCTGAAGCTGACCGCGCGGACCGCCCGCCCAAGGGC 1984

Qy 354 -----AlaProAlaSer-----ThrSerAlaSerValThrLeuValAsnAspLeuAsp 369
Db 1985 AAGCGGCGCCCGGGGCTTCTTCCAGCTGGGCGCCACACG----- 2026
Qy 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2027 -----GTGACCGTCCCGGGCGGGCAGC---GCCTCCGTGACATGACCGCC----- 2071
Qy 390 AspAsnAsnTrpAspGlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGln 408
Db 2072 GACACCCCGCTCGGGCGGCACGCTCGGACGCGCGGTACTCGGGGTACGTGGTCTGCC---ACG 2128
Qy 409 SerGlyThrTyrThrVal-----GluValGlnAlaTyrAsn 420
Db 2129 GCGCGGGGCGACAGCTCCGACGCGCGCGGTGACAGCGGAGGTCTGAGTCTGACGAC 2188
Qy 421 Val-ProGlnGlyProGlnAla 427
Db 2189 GTGACCGTCCGGCACATCGGCC 2210

Search completed: April 4, 2004, 12:03:34
Job time : 98.5977 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 328.004 Seconds

(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-6

Perfect score: 2250

Sequence: 1 NDVARGIVRADVAQSSYGLY.....EVQAYNVPGQAFSLAIYN 434

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapop 6.0			7.0
Delop 6.0			7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO.spool/US09985689/runat_31032004_161809_4271/app_query.fasta_1.3498
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09985689 -CGN 1 1 601 -runat 31032004_161809_4271
-NCPU=6 -ICPU=3 -NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100
-LOGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Published Applications NA:

	1:	/cgn2_6/ptodata1/pubpna/US07_PUBCOMB.seq:
	2:	/cgn2_6/ptodata1/pubpna/PCT_NEW_PUB.seq:
	3:	/cgn2_6/ptodata1/pubpna/US06_NEW_PUB.seq:
	4:	/cgn2_6/ptodata1/pubpna/US06_PUBCOMB.seq:
	5:	/cgn2_6/ptodata1/pubpna/US07_NEW_PUB.seq:
	6:	/cgn2_6/ptodata1/pubpna/PCTUS_PUBCOMB.seq:
	7:	/cgn2_6/ptodata1/pubpna/US08_NEW_PUB.seq:
	8:	/cgn2_6/ptodata1/pubpna/US08_PUBCOMB.seq:
	9:	/cgn2_6/ptodata1/pubpna/US09_PUBCOMB.seq:
	10:	/cgn2_6/ptodata1/pubpna/US09B_PUBCOMB.seq:
	11:	/cgn2_6/ptodata1/pubpna/US09C_PUBCOMB.seq:
	12:	/cgn2_6/ptodata1/pubpna/US09_NEW_PUB.seq:
	13:	/cgn2_6/ptodata1/pubpna/US10A_PUBCOMB.seq:
	14:	/cgn2_6/ptodata1/pubpna/US10B_PUBCOMB.seq:
	15:	/cgn2_6/ptodata1/pubpna/US10C_PUBCOMB.seq:
	16:	/cgn2_6/ptodata1/pubpna/US10D_PUBCOMB.seq:
	17:	/cgn2_6/ptodata1/pubpna/US10E_PUBCOMB.seq:
	18:	/cgn2_6/ptodata1/pubpna/US10F_PUBCOMB.seq:
	19:	/cgn2_6/ptodata1/pubpna/US10G_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

	1	2143	95.2	1305	15	US-10-385-662-1	Sequence 1, Appli
	2	450.5	20.0	1977	13	US-10-090-624-11	Sequence 11, Appli
	3	406.5	18.1	1236	13	US-10-090-624-2	Sequence 2, Appli
	4	406.5	18.1	1962	13	US-10-090-624-15	Sequence 5701, Ap
	5	362	16.1	3624	14	US-10-156-761-5701	Sequence 1, Appli
	6	362	16.1	902568	14	US-10-156-761-1	Sequence 306, Ap
	7	355.5	15.8	3417	14	US-10-156-761-3306	Sequence 1, Appli
	8	355.5	15.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
	9	339	15.1	135638	14	US-10-314-657-1	Sequence 1334, Ap
	10	337.5	15.0	1329	9	US-09-974-300-1934	Sequence 5, Appli
	11	317	14.1	4765	13	US-10-090-624-5	Sequence 5, Appli
	12	297.5	13.2	1560	15	US-10-084-8468-113	Sequence 1, Appli
	13	297.5	13.2	59816	15	US-10-084-8468-1	Sequence 2, Appli
	14	297.5	13.2	59816	15	US-10-084-8468-2	Sequence 33, Appli
	15	284.5	12.6	3788	10	US-09-927-827-33	Sequence 29, Appli
	16	272	12.2	3743	10	US-09-927-827-29	Sequence 1, Appli
	17	272	12.1	1306	9	US-09-966-921A-1	Sequence 5, Appli
	18	272	12.1	1330	14	US-09-966-921A-5	Sequence 5384, Ap
	19	266	11.8	3303	14	US-10-156-761-5384	Sequence 1, Appli
	20	264	11.7	840	14	US-10-209-812-1	Sequence 1, Appli
	21	264	11.7	840	14	US-10-209-812-1	Sequence 112429
	22	261	11.6	1971	9	US-09-974-300-1935	Sequence 1335, Ap
	23	260.5	11.6	2166	12	US-10-344-231-17	Sequence 17, Appli
	24	260.5	11.6	2166	12	US-10-363-332A-17	Sequence 17, Appli
	25	258	11.5	1485	9	US-09-974-300-1938	Sequence 1338, Ap
	26	257	11.4	1140	8	US-08-322-678-11	Sequence 11, Appli
	27	257	11.4	1140	16	US-10-323-324-11	Sequence 11, Appli
	28	257	11.4	1143	14	US-10-313-853-6	Sequence 6, Appli
	29	257	11.4	2588	12	US-10-344-231-20	Sequence 20, Appli
	30	257	11.4	2588	12	US-10-363-332A-20	Sequence 20, Appli
	31	256	11.4	1140	8	US-08-322-678-12	Sequence 12, Appli
	32	256	11.4	1140	16	US-10-323-324-12	Sequence 12, Appli
	33	250.5	11.1	1140	15	US-10-146-905A-9	Sequence 9, Appli
	34	246.5	11.0	1140	9	US-09-920-118-13	Sequence 13, Appli
	35	244.5	10.9	1497	8	US-08-322-678-6	Sequence 6, Appli
	36	244.5	10.9	1497	9	US-09-060-854B-1	Sequence 1, Appli
	37	244.5	10.9	1497	14	US-10-033-325-1	Sequence 1, Appli
	38	244.5	10.9	1497	14	US-10-228-572-1	Sequence 1, Appli
	39	244.5	10.9	1497	15	US-10-423-649-1	Sequence 6, Appli
	40	244.5	10.9	1497	16	US-10-323-324-6	Sequence 18, Appli
	41	244	10.8	2267	12	US-10-344-231-18	Sequence 18, Appli
	42	244	10.8	2267	12	US-10-363-332A-18	Sequence 18, Appli
	43	242.5	10.8	1149	15	US-10-146-905A-7	Sequence 7, Appli
	44	242.5	10.8	1494	14	US-10-104-693-1	Sequence 14, Appli
	45	240.5	10.7	1074	9	US-09-837-235-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

;; PRIOR FILING DATE: 2002-10-18
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 1
;; LENGTH: 1305
;; TYPE: DNA
;; ORGANISM: Bacillus sp. KSM-KP43
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1305)
;; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 7,45e-226 Length: 1305
Score: 2143.00 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 9
Query Match: 95.24% Indels: 0
DB: 15 Gaps: 0

US-09-985-689a-6 (1-434) x US-10-385-662-1 (1-1305)

Qy 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTTGGCGGTGGAAATTTGTCAAAGCGGATGTGGCTCAGACGACTACGGGTTCAT 60
Qy 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAAGGACACATGCTAGCGGTGGCGATACAGGGCTTGATACAGTGGCAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGTCATGAAGCTTCGCGGGGAAAAATTACTGCATTATATGCAATGGGACGGACGAAT 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCCAATGATACCAATGCTCATGCTACGATGCGTGGCTCGCTATTAAGGAAAGCGC 240
Qy 81 ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSer 100
Db 241 TCCACTATAAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTATCATGATGATG 300
Qy 101 AsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSer 120
Db 301 GTGGGGGACTTGGAGGACTCTTCGAATCTGCAAACTTATTTCAGCCAGCATACAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 140
Db 361 GCTGTGTCAGAAATTCATAAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 420
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAla 160
Db 421 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 480
Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 CGGAATGAGGACCGAACCGCGGNAACCATCAGTCACGACGACGACGATAAATAATGCAATA 540
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTCGGAGCTACGGGAAACCTCCGCCCAAGCTTGGTCTTATGCGGACAAATATCAAC 600
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 601 CATGTGGCAGATCTCTTCATCGTGGACCGACAAAGGATGGACGATCAAAACCGGATGTC 660
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 661 ATGGCACCGGAAACGTCATCTATCATCAGCAAGATCTTCTCTTGGACCGGATTCCTCCTTC 720
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260

Db 721 TGGCGGAACCATGACAGTAATATGTCATATGGGTGGAAAGCTCCATGGCTTACACCGATC 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro 280
Db 781 GTTGCTGGAAACGTGGACAGCTTCGTGAGCATTTTGTGAAAAACACAGAGGATCACACCA 840
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyr 300
Db 841 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCCGTGCAGCTGACATCGGCTTGGCTTAC 900
Qy 301 ProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 901 CCGAACGGTAACCAAGGATGGGACGATGACATTTGGTAAATCCCTGAACGTTGGCTTAC 960
Qy 321 ValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAlaGln 340
Db 961 GTGAACGAGTCCAGTCTCTATCCACAGCCAAAGAGCGAGCTACTTCGTCTTACTGCTACT 1020
Qy 341 SerGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrSerAla 360
Db 1021 GCGGCAAGCCTTTGANAATCTCCCTGTATGGTCTGATGCTCCCTTCGAGGACACACTGCT 1080
Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysTyr 380
Db 1081 TCGGTAAACGTTGTCAATGATCTGGACCTTGTCAATTCAGCTCCAAATGGCACACAGTAT 1140
Qy 381 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1141 GTAGGAATGACTTTACTTCGCCATACATGATAATCGGATGGCGGCAATTAACGTAGAA 1200
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 420
Db 1201 AATGATTTATTAATGCCACCAAGCGGAGCTATACAAATGAGGTACAGGCTTATAAC 1260
Qy 421 ValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1261 GTACCGGTTGGACCAACAGACCTTCTCGTTGGCAATTTGTAAT 1302

RESULT 2

US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 3,14e-39 Length: 1977
Score: 450.50 Matches: 138
Percent Similarity: 47.05% Conservative: 69
Best Local Similarity: 31.36% Mismatches: 170
Query Match: 20.02% Indels: 63
DB: 13 Gaps: 15

```
US-09-985-689A-6 (1-434) x US-10-090-624-11 (1-1977)
QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnValValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACTCCCTCGCTACGAGGAGCGGTCTGTGTGTC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTCGTATACGGGTATACAGCGAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaIleTyr---AlaLeuGlyAlaGlnSerValMet-----AspSerAsn 66
Db 535 GGCAGGTCATAGGTCGTACAGCGCGTCAACGCGGAGTCCGACCCCTACGATCAGCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsnLys-----84
Db 595 GGACAGGAGAACACCGTTCGGGTATCGTTCGGGAACCGGAGCGGTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet-----AspSerAsn 101
Db 655 ATAGGGTCGCCCCCGCGGAGCTGTCGGGTCAGAGTTCTCGTGCAGCGGTTCTG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGGGGTGTGACTGGTCTCCAGAACAGGACAGTAC 774
QY 122 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAsp 141
Db 775 GGATAGGGTATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGAGCGGACCGACTCC 834
QY 142 SerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGly 161
Db 835 CTCAGTCAGCGCGTCAACACGCGCTGGGAGCGGTATAGTAGTCTCGGTCGCGCGGC 894
QY 162 AsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 895 AACAGCGGCGGACACCTACACCGTCTCCCGCGCGCGCGGAGCGGAGGTCATAACC 954
QY 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 955 GTCTGTGCA-----GTTGACAGCAACGACAC 981
QY 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
Db 982 ATCGCCAGCTTCTCCAGCAGGAGCGGACCGCGGAGCGGAGGCTCAAGCGGAGTCGTC 1041
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 241
Db 1042 GCCCGCGGCTTACATATAGCCCGCGCGCGCAGC-----GGAACAGCATGGGC 1092
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 261
Db 1093 ACCCGGATAAACGACTACTACACCAAGGCTCTCGGAACGAGATGGCCACCCCGCAGTT 1152
QY 262 AlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrPro---280
Db 1153 TCGGGCGGTGGCGCTCATCTCTCCAG-----GCCACCGGAGTGGACCCCGGAC 1203
QY 281 -----LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIleGly---297
Db 1204 AAGGTGAAGACCGCTCATCGAGACCGCGGACATAGTCGCCCCCAAGGAGATAGCGGAC 1263
QY 298 LeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu---316
Db 1264 ATCGCTAC-----GGTGGGGTAGGGTGAACGTCTACAAGGCCATCAAG 1308
QY 317 -----AsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr 334
Db 1309 TAGACGACATACGCCAAGCTCACCTTCACGGCTCGTCCGCGGACAGGAGGAGCCGAC 1368
QY 335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAspAla 354
Db 335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrpSerAspAla 354
Db 1369 CACACCTTCGACGTACGCGCGCCACCTTCGTTCGTCACCGCCACCTCTACTTGGGAC-----1422
QY 355 ProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAla 374
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTCTACCTCTACGAC 1458
QY 375 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 394
Db 1459 CCCAACCGGAGCAGG---GTTGACTACTCTCTACACCGCTACTAC-----1500
QY 395 GlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrVal 414
Db 1501 -----CGCTTCGAGAGGTCGCTACTACACCGCGCGGACCTGACCGGTC 1551
QY 415 GluValGlnAlaTyrAsnValProGlnGlyProGlnAlaPheSerLeuAlaIleValAsn 434
Db 1552 AAGTCTCTCAGCTAC-----AAGGGCGGCGGAACTTACCAGGTTCGACGTCGTCAGC 1602
RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US20020132395A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR FILING DATE: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2
Alignment Scores:
Pred No.: 1,09e-34 Length: 1236
Score: 406.50 Matches: 136
Percent Similarity: 44.22% Conservative: 63
Best Local Similarity: 30.22% Mismatches: 169
Query Match: 18.07% Indels: 82
DB: 13 Gaps: 17
US-09-985-689A-6 (1-434) x US-10-090-624-2 (1-1236)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCGAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGATATGATGTTCTGGAAT 83
QY 24 nValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACATAGGATATTCACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaIle-- 63
Db 126 AGATCTCAAGGAAAGTA-----ATTGGGTGGGTAGATTGTTCAATGG 170
QY 64 -----AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACATGACCATGACATGACATGACATGACATGACATGACATGACATG 230
QY 78 yAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
```

[illegible]


```

Qy 231 ArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyr 250
Db 1216 CGTCCGAGCTCCCGCCGACCGGCTAC-----TACACCTCC 1254
Qy 251 MetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGlu 270
Db 1255 ATGAGCGGTACGTGATGAGCGCCCGGCGGCGGTGCGCGGCTCTCTCGCGAG 1314
Qy 271 HisPheIleLysAsnArgGlyLeuThrProLysProSerLeuLeuLysAlaAla----- 288
Db 1315 CAGCACCCGACGTGACCGCGCGGCTCAGGACGCGGTGATGTCACGTCGAGCAA 1374
Qy 289 LeuIleAlaGlyAlaThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGly 308
Db 1375 CTCGAGCCTCCGTATATAGTTGGGGCGGCTCGGTGAGTGTCCCGACCGCTCGGC 1434
Qy 309 ---ArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGlnThrSerSerLeu 327
Db 1435 GCCCGGTACCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1494
Qy 328 SerThrAsnGlnLysAlaThrTyrSerPheThrAlaGlnSerGlyLysProLeuLysIle 347
Db 1495 GATCGACCCGCTACGAAAGACGCTCACTACTCCAACTCTCCGACACGAGCTCGAGTTG 1554
Qy 348 SerLeuValTrpSerAspAlaProAlaSerThrSerAlaSerValThrLeuValAsnAsp 367
Db 1555 AGCTTCGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1599
Qy 368 LeuAspLeuValIleThrAlaPro-----AsnGlyThrLysTyrValGlyAsn 383
Db 1600 ---GACACCGCACTACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1656
Qy 384 AspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGluAsnValPhe 403
Db 1657 GGTCCAGGCTCCGCTG-----GGAGACGAGCGGCGGCGGCGGCGGCGGCGGCGG 1698
Qy 404 IleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnValProGln 423
Db 1699 -----GCCCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1749
Qy 424 GlyProGlnAlaPheSerLeuAlaIle 432
Db 1750 GAGCGGAGCGGTACTCCCTCACCNGT 1776

```

RESULT 6

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCES: 249-262

; CURRENT APPLICATION NUMBER: US/10/156, 761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.:	4,86e-24	Length:	9025608
Score:	362.00	Matches:	135
Percent Similarity:	43.88%	Conservative:	62
Best Local Similarity:	30.07%	Mismatches:	186
Query Match:	16.09%	Indels:	66
DB:	14	Gaps:	17

US-09-985-689A-6 (1-434) x US-10-156-761-1 (1-9025608)

```

Qy 8 ValLysAlaAspValAlaGlnSerSer-----Tyr 17
Db 6918813 GTCGAGGCGGACATGCGCGAGAGCAACGCGCAGATCGGTACGCGGCGCGGTGGAGGCC 6918872
Qy 18 GlyLeuTyrGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 6918873 GGGCTCACGGGCGAGCGGCTCACCGTCCGCGTGTGCACACCGCGCGTCGACATC----- 6918926
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGly 57
Db 6918927 -----ACTCACCCCGACCTCGCGCGCGGCGGTGCCGAGCAAGAGCTTCATC 6918974
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 6918975 GACGGGAGGAGGTGCGCGACCGCAACGCGCCACGCGACCCACGTCACCTCGACCGCTCGGC 6919034
Qy 78 GlyAsnGlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 6919035 GCACGCGGCGCGCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919094
Qy 94 PheGlnSerValMetAspSerAsnGlyGlyLeuGlyLeuProSerAsnValSerThr 113
Db 6919095 GTCGCAAGTGCTC---AGCGACGAGGCGCGGGAAGCGAGTCCAGATCATCGCGGCG 6919151
Qy 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaPro 133
Db 6919152 ATGGAATGGCGCGCGGCGGACGTGCTGCCAGGATCGTCTCGATGASCTCGGATCG--- 6919208
Qy 134 ValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAspAspTyrValArgLys 151
Db 6919209 ACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919268
Qy 152 AsnAspMetAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSer 171
Db 6919269 ACCGCGCGCTCTTCGTCGTCGCGCGGAAACACCGGTGCC---CCCTCTCGATCGGC 6919325
Qy 172 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSer 191
Db 6919326 TCGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919364
Qy 192 PheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThr 211
Db 6919365 -----GTGACTCATCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919412
Qy 212 Lys---AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAla 230
Db 6919413 CACGGCGCAACGCGCTCAAGCCGACCTCCGCCACCCGCGCTCGACATCCCGCGCGCGCC 6919472
Qy 231 ArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyr 250
Db 6919473 CGCTCCAGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919511
Qy 251 MetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGlu 270
Db 6919512 ATGAGCGGTACGTGATGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919571
Qy 271 HisPheIleLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAla----- 288
Db 6919572 CAGCACCCGACGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6919631

```


US-09-974-300-1934

Alignment Scores:
Pred. No.: 4,81e-27 Length: 1329
Score: 337.50 Matches: 108
Percent Similarity: 45.18% Conservative: 42
Best Local Similarity: 32.53% Mismatches: 115
Query Match: 15.00% Indels: 67
DB: 9 Gaps: 13

US-09-985-689A-6 (1-434) x US-09-974-300-1934 (1-1329)

QY 11 AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyValValAlaValAlaAsp 30
DB 406 GAATGTGTGAGAAACATCAGACCTGACAGGCAAGAGTGCAGCTGCTGTCATGAT 465
QY 31 ThrGlyLeuAspThrGlyArgAspSerSerMetHisGluAlaPheArgGlyLysIle 50
DB 466 ACGGCGGTA-----TACCCTCAGCAAGATCTTGAAGCAGGATC 504
QY 51 ThrAlaIleTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGly 69
DB 505 AGGCTTTTCAAGACTTATCAACAGAGAAACAGACCTATGATGACATGGCAGCGC 564
QY 70 ThrHisValAlaGlySerValLeuGlyAsnGlyThrSerAsn-----LysGly 85
DB 565 ACACACTCGCGCGGTGATGCTTGGGAAACGAGCGGCTCATCGGTCAGTACCGCGGA 624
QY 86 MetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGlyGlyLeuGly 105
DB 625 CCTGCTCTGAGCAGAACTTGTGGTGTAAAGTATTTGGACAAATGGGATCC---GGA 681
QY 106 GlyLeu-----Pro 108
DB 682 TCGCTCGAAACCGTCATTCAGGCGTACATTTGGTGTCAATTCATTAAGGAAATCCT 741
QY 109 SerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn 128
DB 742 GATGATCCGATCGACATATTATTCATGATCTGGTGCAGAGCCTTGGCTACGAGAT 801
QY 129 SerTrpGlyAlaPro-----ValAsnGlyAlaTyrThrThrAspSerArgAsn 144
DB 802 GAAGAAGAAGATCCAGTCGTTAAAGCTGTTCAATGATGCGAGCATGGGACGCA----- 849
QY 145 ValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGlyAsnGlyGly 164
DB 850 -----GGCATTTGTTGTATGTCGGCAGCGCGGCAACTCCCGT 885
QY 165 ProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 184
DB 886 CTGATGCCCAACGATTCGCCAGCGCGGTGTCAGCAGCAAGATTTATTACAGTCGGAGC 945
QY 185 ThrGluAsnLeuArgSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 204
DB 946 TTGGATGAC---AGGATACAGTCAGCGCGGAGGATGACGAT-----GTGCGCTCT 993
QY 205 PheSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGly 224
DB 994 TATTCAAGCAGAGGCGGCAATCTATGCTCAAGTCRAACCGCATTTGCTGTACCGGCGC 1053
QY 225 ThrPheIleLeuSerAlaArgSer-----SerLeuAlaProAspSerSerPheTrpAla 242
DB 1054 ACAATATTACGTCGCTTCGTTCCACCGGATCTTTTCTCGATAAGCTGCAAAACAAAC 1113
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 262
DB 1114 AGAGTCGGCACAAAATATATGACATTCGCGAACCTCGATGCTACCGCGATCTCGCGCA 1173
QY 263 GlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysPro 282
DB 1174 GGAATTGGCGCA-----CTTATCTTCAGCAAGCGCGGCGCACAGAACT 1218
QY 283 SerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThrAspIle----- 296

US-09-985-689A-6 (1-434) x US-10-090-624-5 (1-4765)

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1024 GGCAATGGTTATGACATTCGATATGATGATGATGATGATGATGATGATGATGATGAT 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
DB 1084 GTTCCACTTGGCCAGTACAACTTACTATGATGATGATGATGATGATGATGATGATGATGAT 1143
QY 49 LysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn----- 66
DB 1144 CCTCTCAACTACGTGCTTGA-----GAATAGATCTTACCGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
DB 1186 TATGCGATATTGGTGGGATGTCAGGTCACCGAACTCACGTAAGTCTGTAAGTCTGTAAGT 1245
QY 78 GlyAsnGlyThrSerAsn----- 83
DB 1246 GGTTCACAGCAACAATGATGCTGGGATGCTGGGATGCTGGGATGCTGGGATGCTGGGAT 1305
QY 84 -----LysGly 85
DB 1306 GTGTTCTCAAGACTTATGTTGGGATTTACCAAGCTTACCAAGCTTACCAAGCTTACCAAG 1365
QY 86 MetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGlyGlyLeuGly 105
DB 1366 GTTCTCCAGGTGCGCAATATAGGCAATATAGGCAATATAGGCAATATAGGCAATATAGGCA 1419

US-10-090-624-5

Sequence 5, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hi-karu
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyo-ko
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA=6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 4765
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:
Pred. No.: 5,82e-24 Length: 4765
Score: 317.00 Matches: 131
Percent Similarity: 39.87% Conservative: 50
Best Local Similarity: 28.85% Mismatches: 142
Query Match: 14.09% Indels: 131
DB: 13 Gaps: 15

QY 106 GlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle 125
Db 1420 GGTAGCATGGGATATTATAGAGGATGATGACATACGCAACCCATGGTGAGACGTT 1479
QY 126 HisThrAsnSerTyrGly-----AlaProValAsnGlyAlaTyrThrAspSerArg 143
Db 1480 ATAAGCATGAGTCTCGTGGGAATGCTCCATCTAGATGGTACTGATCCAGAAAGCGTT 1539
QY 144 AsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGlyAsnGlu 163
Db 1540 GCTGTGATGATGCTTACGAAAGTACGGTGTGTTATTCGTAATAGTGTGAGGAATGAA 1599
QY 164 GlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 183
Db 1500 GGTCTCGCATTAACATCGTTGGAAAGTCTGGTGTGTCACAAAGGCAATACCTGTGGA 1659
QY 184 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsp----- 197
Db 1660 GCTGCTGCA---GTGCCCATTAACGTGGAGTTATGTTCCCAAGCACTGGATATCCT 1716
QY 198 -----AsnIleAsnHisValAlaGlnPhe 205
Db 1717 GATTACTGATGATCTATTACTCCCGCGCTACACAAACGTT---AGATAGCATTTCTTC 1773
QY 206 SerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThr 225
Db 1774 TCAAGCAGAGCGCGAGATAGATGTTGAATAAACCCTATGATGGTCTCCAGGTAC 1833
QY 226 PheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAsp 245
Db 1834 GGAATTTACTCATCTCCCGCGATGTTGATGGAGCTGAGCTGCTC----- 1878
QY 246 SerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnVal 265
Db 1879 -----ATGCTCGNACTTCGATGGTACTCTCATGCTCAGCGTGTGCTT 1923
QY 266 AlaGlnLeuArgGluHisPheIleLysAsnArgGlyIleThrProLysProSerLeuLeu 285
Db 1924 GCATCTCTCATAGCGGG---GCAAGCGCGAGGGAATATACTACATCCAGATATAAAT 1980
QY 286 LysAlaAlaLeuIleAlaGlyAlaThrAspIle-----GlyLeuGly 299
Db 1981 AAGAAAGTCTTTGAGCGGTGCAACTGCTGGTTCGAGGAGATCCATATATCGGCGAAG 2040
QY 300 TyrProSerGlyAsnGlnTyrTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 319
Db 2041 TACACTGAGCTTGACCAAGGTCATGGTCTTTAAAGTTACCAAGTCTCGGAAATC--- 2097
QY 320 PheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThrTyrSerPheThrAla 339
Db 2098 -----CTTAAGGCT 2106
QY 340 GlnSerGlyLysProLeuLysIleSerLeuValTrpSerAsp-AlaProAlaSerThrSe 359
Db 2107 ATAAACGGCACCATCTCCCAATTTGATCACTGGCGACAGCAAGTCTTACAGGACTTT 2166
QY 359 rAlaSerValThrLeuValAsnAspLeu-----AspLeuValIle 372
Db 2167 GCGGAGT---ACTTGGGTGTGGAGCTTATAAGAGGTCTCTACGCAAGGAATCTATACCT 2223
QY 372 eThrAlaProAsnGlyThr--LysTyrValGlyAsn 383
Db 2224 GACATTGTGAGTGGCATTAAAGTACGTAGGGGAC 2259

RESULT 12

US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLERWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS

; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084.846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avx16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:

Pred. No.:	1-55e-22	Length:	1560
Score:	297.50	Matches:	120
Percent Similarity:	41.36%	Conservatives:	50
Best Local Similarity:	29.20%	Mismatches:	166
Query Match:	13.22%	Indels:	75
DB:	15	Gaps:	19

US-09-985-689A-6 (1-434) x US-10-084-846A-113 (1-1560)

QY	21	GlyGlnGlyGlnValValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	451	GGGCAGGAGTGACGGCTAGCTATCATGCACACCGCGCTC-----CGC	492
QY	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyr---AlaLeuGlyArgThr	59
Db	493	ATCACCCACAGGACTTCGGCGCGGCGCTCTACGCTACGACGCCATCGACACACGAC	552
QY	60	AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn	79
Db	553	AAACCGGCCAGGACGGCCAGCGCACGCGCACGCGTGGCGGCACGCTCGCGGCAAC	612
QY	80	GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp	99
Db	613	GCC-----TACGGGTGCCAAGAGGCCAAGATCTGATGGGTCCGGTGTCTG--- 660	
QY	100	SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyr	119
Db	661	AAACAACCTCGGCGCAGGCGACCCAGGTCGTCGCGCGCATCGACTGGTTCGCCCGG	720
QY	120	SerAlaGlyAlaArgIleHisThrAsn---SerTrpGlyAlaProValAsnGlyValTyr	138
Db	721	AAAGCGCTCAAGCGCGCGCTCGCCCAACATGCTCTCGCGCGCGCGCGCACGCGCTC	780
QY	139	ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPhe	158
Db	781	GACACGGCGGTAGGCAAC-----GCCATGGCTCCGGGCTCACCTTCGCGGTG	828
QY	159	AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn	178
Db	829	GGCGCGCGCAACAGCTCGACCAACGCTCCACAGGTCAC---CCGCGCGCTCACCGAG	885
QY	179	AlaIleThrValGlyValThr-----GluAsnLeuArgProSerPheGlySerTyrAla	196
Db	886	GCAATCAGGTGCGCGCGCGACGACGACGAGTCGACGCGCAAGCGCGGTACTTCCACT	945
QY	197	AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIle	216
Db	946	TCCGTCTCTC----- 954	
QY	217	LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro	236
Db	955	-----GACCTCTTCGCGCGCGGTCTTCCTCATCACTCGGCC----- 990	

QY 237 AspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMet 256
 Db 991 TGAACCTCAAGCGACTCGCGACCAACACCATCTCCGGTACGTCCGATG 1038
 QY 257 AlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArg 276
 Db 1039 GCGACCCCGACGCGGGGGCGCGCGCGTC-----CACCTCGCGCGCAACCCC 1089
 QY 277 GlyIleThrProLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
 Db 1090 TCGGCCACCCCGTCCAGGTCCGACGCGGTGACGTCGCGCGCACCAACCGCGCGTCGTC 1149
 QY 294 ThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAsp 313
 Db 1150 ACCAACCCCGCACGCGGTCCCAACCGGCTCCTGTACGTGCGCGCG-----CGCGACGAC 1206
 QY 314 LysSerLeuAsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAla 333
 Db 1207 CACCTCCGCGCGCGCGTTCGAGAAC---ACCGGTGACTACACGATCAGCGACCACTCC 1263
 QY 334 ThrTyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValThrSerAsp 353
 Db 1264 ACGETTCGAG-----TCCCGGTGACGGTCTCCGCGCTCCGCGCAAC 1305
 QY 354 AlaProAlaSerThrSerAlaSerVal-----ThrLeuValAsnAspLeuAsp 369
 Db 1306 GCGCCCTCGGCCCTCGGTAGAGTCCACATCGTCCACAGTACATACGCGGACCTCCAG 1365
 QY 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
 Db 1366 GTCCAGCTGATCGCCCGCGGACGCGCGGTACAGCTCAAGTCTGATC-----1413
 QY 390 AspAsnAsnTyrAspGlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln 408
 Db 1414 ---GGACCGCGCGGAGTTGGACCAACATCAACACCGTACTCGGTGACGCTCTCTCG 1470
 QY 409 -----SexGlyThrTyrThrValGluVal 416
 Db 1471 GAGGCGGCGCACCGGACGTTGGAACACTCGCGGTG 1503

RESULT 13

US-10-084-846A-1
 ; Sequence 1, Application US/10084846A
 ; Publication NO. US2004006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MUEHLENWEG, AGNES
 ; APPLICANT: TREPFER, AXEL
 ; APPLICANT: BECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084,846A
 ; PRIOR FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: PatentIn Ver. 3.2
 ; SEQ ID NO 1
 ; LENGTH: 59816
 ; TYPE: DNA
 ; ORGANISM: Streptomyces viridochromogenes
 US-10-084-846A-1

Alignment Scores:
 Pred. No.: 3,52e-20 Length: 59816
 Score: 297.50 Matches: 120
 Percent Similarity: 41.36% Conservative: 50
 Best Local Similarity: 29.20% Mismatches: 166
 Query Match: 13.22% Indels: 75
 DB: 19 Gaps: 15

US-09-985-689A-6 (1-434) x US-10-084-846A-1 (1-59816)
 QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 56648 GCGCAGGAGTGCAGCGCGTACGTATCATGCACACCGCGCGTC-----CGC 56689
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyr---AlaLeuGlyArgThr 59
 Db 56690 ATCACCACAGGAGCTTCGGCGCGCGCGCTCTCTACGGCTACGACGCCATCGACACACGAC 56749
 QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 56750 AACACCGCGCGACG 56809
 QY 80 GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp 99
 Db 56810 GCC-----TAGCGGTCCCAAGAGGCGCAAGATCGTAGCGGTCCGCGTCTGCTG--- 56857
 QY 100 SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyr 119
 Db 56858 AACAACTCCGGCG 56917
 QY 120 SerAlaGlyAlaArgIleHisThrAsn---SerTyrGlyAlaProValAsnGlyAlaTyr 138
 Db 56918 AACGGCGTCAACCG 56977
 QY 139 ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPhe 158
 Db 56978 GACACGCGCGTACGCAAC-----GCCATGGCGCGCGGTCACTTCGCGCGTG 57025
 QY 159 AlaAlaGlyAsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsn 178
 Db 57026 GCGCGCGCGACGAGTCGACCAACGCTCCACGAGGTCA---CCGCGCGCGTCCCGGAG 57082
 QY 179 AlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGlySerTyrAla 196
 Db 57083 GCCATCACGTCG 57142
 QY 197 AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIle 216
 Db 57143 TCGGTCTCT-----57151
 QY 217 LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPro 236
 Db 57152 -----GACTCTTCGCGCGCGGTTCGTCCATCACCTCGCGC-----57187
 QY 237 AspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMet 256
 Db 57188 -----TGAACCTCAAGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 57235
 QY 257 AlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLysAsnArg 276
 Db 57236 GCGACCCCGCACGTCG 57286
 QY 277 GlyIleThrProLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
 Db 57287 TCGGCCACCCCGTCCCG 57346
 QY 294 ThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyrGlyArgValThrLeuAsp 313
 Db 57347 ACCAACCCCGCACGCGGTCCCAACCGGCTCCTGTACGTGCGCGCG-----CGCGACGAC 57403
 QY 314 LysSerLeuAsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAla 333
 Db 57404 CACCTCCGCGCGCGCGTTCGAGAAC---ACCGGTGACTACACGATCAGCGCAACACTCC 57460
 QY 334 ThrTyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValThrSerAsp 353
 Db 57461 ACGGTCCGAG-----TCCCGGTGACGCTCCCGCGCGTCTCCGCGCAAC 57502
 QY 354 AlaProAlaSerThrSerAlaSerVal-----ThrLeuValAsnAspLeuAsp 369
 Db 57503 GCGCCCTCGCGCGCTCGCGGTAGAGGTCCACATCGTCCACACGATACATCGCGGACCTCCAG 57562

QY 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 57563 GTCCAGCTGATCCGCCCGGACGCGGCGTACACGCTCAAGTCGTAC----- 57610
QY 390 AspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln 408
Db 57611 ---GGCACCGCGCGAGTTCGGGACACATCAACACCATCTACTCGGTGAACGCTCCTCG 57667
QY 409 -----SerGlyThrTyrThrValGluVal 416
Db 57668 GAGCGCGCCACGCGACGTGGAACTGCGGGTG 57700

RESULT 14

US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLERWEG, AGNES
; APPLICANT: TREFFER, AXEL
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes

US-10-084-846A-2

Alignment Scores:

Pred. No.: 3 526-20 Length: 59816
Score: 297.50 Matches: 120
Percent Similarity: 41.36% Conservative: 50
Best Local Similarity: 29.20% Mismatches: 166
Query Match: 13.22% Indels: 75
DB: 15 Gaps: 19

US-09-985-689A-6 (1-434) x US-10-084-846A-2 (1-59816)

QY 21 GlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGGAGTGACGGGTACGTATCGACACCGGGCTC-----CGC 3128
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyr---AlaLeuGlyArgThr 59
Db 3127 ATCACCACAGCAGCTTCGGCGCGCGGCTCTCTACGGCTACGACGCCATCGACACAGC 3068
QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 3067 AACACCCCGCAGGACGCGCCACGCGCACGCGACGCGTGGCGCGGACGCTCGCGCGCAAC 3008
QY 80 GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp 99
Db 3007 GCC-----TACGGCGTCGCCAAGAGGCCAAGTCTAGCGTCCCGCTGCTGCTG--- 2960
QY 100 SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyr 119
Db 2959 AACAACTCCGGCCAGGCGCACCGCGCCAGCTGCTCGCGGCTACGCTGGTGGCGCGG 2900
QY 120 SerNlaGlyAlaArgIleHisThrAsn---SerTrpGlyAlaProValAsnGlyAlaTyr 138
Db 2899 AACGCCGTGAAGCGCGCGCTGGCCAAATGTCCTCTCGCGCGCGCGCGCGCGCTC 2840
QY 139 ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPhe 158

RESULT 15

US-09-927-827-33

; Sequence 33, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927, 827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69

Db 2839 GACACGGCGCTACGCAAC-----GCCATGGCTCCGGCGTCACTTCGGCGTG 2792
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
Db 2791 GCGGCGCGCAACGAGTCGACCAACGCTCCACGAGTCA---CCGCGACGCTCACCGAG 2735
QY 179 AlalleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGlySerTyrAla 196
Db 2734 GCCATCAGCTCGCGCGCGACGACGAGTCGAGCCCAAGCCGCTACTCCAACTACGCGC 2675
QY 197 AspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIle 216
Db 2674 TCCGTCTC----- 2666
QY 217 LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaPro 236
Db 2665 -----GACCTCTTCGCGCGCGGTTCGTCCATCACTTCGGCGC----- 2630
QY 237 AspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMet 256
Db 2629 -----TGAACCTCAAGCGACTCGCGGCGCGCGCTC-----CACCTCGCGCGCAACCC 2582
QY 257 AlaThrProIleValAlaGlyAsnValAlaGlnLeuArgLysPheIleLysAsnArg 276
Db 2581 GCGACCCCGCACGCTGGCGGCGCGCGCTC-----CACCTCGCGCGCAACCC 2531
QY 277 GlyIleThrProLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
Db 2530 TCGGCCACCCCGCTCCAGGTCCGACGCGCTGAGTCCGCGCGCACACCGCGGTGTC 2471
QY 294 ThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 313
Db 2470 ACCAACCCCGCACGCGGTGCGCCCAACCGGCTCTGTAGCTCGCGCGG---CGGACCGAC 2414
QY 314 LysSerLeuAsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAla 333
Db 2413 CACCTTCGGCGCGCGCTTCGAGAAC---ACCGGTGACTACGATCAGTACGACGCACTCC 2357
QY 334 ThrTyrSerPheThrAlaGlnSerGlyLysProLysIleSerLeuValTrpSerAsp 353
Db 2356 ACGGTGCGAG-----TCCCGGTGACGGTCTCCGGGTCTCCGGCAAC 2315
QY 354 AlaProAlaSerThrSerAlaSerVal-----ThrLeuValAsnAspLeuAsp 369
Db 2314 GCGGCTTCGGCGCTCGCGTAGAGTCCACATCGTCACACGATCAGTACGCGGCGCTCCAG 2255
QY 370 LeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyr 389
Db 2254 GTCCAGCTGATCGCCCGCGCGCGCGGTACGCTCAGTCAAGTCGTAC----- 2207
QY 390 AspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln 408
Db 2206 ---GGCACCGCGCGCGAGTTCGGAACAACATCAACACGAGTACTCGGTGACGCGCTCCTCG 2150
QY 409 -----SerGlyThrTyrThrValGluVal 416
Db 2149 GAGCGCGCAACGCGCACGTGGAACCTCGGGTG 2117

```

; SEQ ID NO 33
; LENGTH: 3788
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2788)
US-09-927-827-33

Alignment Scores:
Pred. No.: 1,57e-20 Length: 3788
Score: 284.50 Matches: 114
Percent Similarity: 40.76% Conservative: 47
Best Local Similarity: 28.86% Mismatches: 112
Query Match: 12.64% Indels: 122
DB: 10 Gaps: 18

US-09-985-689a-6 (1-434) X US-09-927-827-33 (1-3788)
QY 3 ValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGln 22
DB 1739 GTCTGGGCAACGTGCTGTCATGTGGATGACGACGCGCGATATCGGCACGTGGGCAG 1798
QY 23 GlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMet 42
DB 1799 CAGATCGGCTGGCGGCTCTGGATACCGGCATT-----GCGGCGCG 1840
QY 43 HisGluAlaPhe-----ArgGlyLysIleThrAlaIleTyr----- 54
DB 1841 CATCGCATTTCTTCGTCGCGGCGAGCGGACAAATGTGGGCGCAATGGGATGTCAGC 1900
QY 55 -----AlaLeuGlyArgThrAsnAlaAsnAsp 64
DB 1901 CGGCGGCGGCGCCCAAGCGGTGACGCGGCGCGATGCGGCGCTGTTCGCGGCGCGCGTG 2020
QY 65 ProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
DB 1961 GGGCACGGCATGGCACGACATCGCGCGCATCATCGCGCGTTCACCGGTCACCGACTCGAC 1960
QY 79 -----AsnGlyThrSerAsnLys-----GlyMetAlaProGlnAla 90
DB 2021 CTGTATGCCAGTGGGAAGCGGGGGAAGTGTGGGAATTTGCGGATGCGCGCGGAGACG 2080
QY 91 AsnLeuValPheGlnSerValMetAspSerAsnGlyLeuGlyGlyLeuProSer--- 109
DB 2081 CAACTCTATGGCTTCAAGTGTCTGGACGACGCGCGC-----AACGGCGCGGATTTCGTGG 2134
QY 110 -----AsnValSerThrLeuPheSerGlnAla----- 118
DB 2135 ATGATCAAGGCAAGTGGACGCGTGGCGGATCTCAACGAGCGTGGCGGAGCTGGTGATC 2194
QY 119 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProVal 134
DB 2195 CACGGCGTCAATCTCAGCTGGCGGCTACTTCGATCGGAAAGCTACGGCTGTGGCTTC 2254
QY 135 AsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMet 154
DB 2255 -----ACGCGGTTATGCAACAGAGTTGCGCGGCTTATGGCGGCGGCGGCGTA 2299
QY 155 AlaValLeuPheAlaAlaGlyAsnGlyGly-----ProAsnGlyGly 168
DB 2300 CTGGTAGTGGTGGCGCGGCAACGAAAGCGCTGGCTGGCTGATGCGAAGCGCGCGC 2359
QY 169 Thr-----IleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 2360 ACCTATCCGCGCAACATGGATCTGTGATCAGCATCCGGGCAATCTGGAGACGCGATC 2419
QY 181 ThrValGlyValaThrGluAsnLeuArgPro---SerPheGlySerTyrAlaAspAsnIle 199
DB 2420 GTGGTGGATCGGTGACACAGACGCGCGCAATTACGCG----- 2461
QY 200 AsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAsp 219

```

Search completed: April 5, 2004, 02:55:48
Job time: 7399 secs

```

DB 2462 -----GTGTCGTATTTTTCATCCGCGGCGGCGGCGGCGGCGCTCCAAACCTGAT 2515
QY 220 ValMetAlaProGlyThrPheIleLeuSerSerAlaArgSerSerLeuAlaProAspSerSer 239
DB 2516 GTGGTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2566
QY 240 PheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro 259
DB 2567 -----CGCGACCGCGTCCAGCCTGATGTCGAGATGAGCGGCAACAGATGCGCGCACCG 2620
QY 260 IleValAlaGly----- 263
DB 2621 CATGTGTCGCGGCTGCTGGCGGGTTTATCCGACGCGCGGAGTTTCATCGGCTTTCCG 2680
QY 264 ---AsnValAlaGlnLeu-----ArgGluHisPheIle 273
DB 2681 GACCGGTCAGCAACTGCTGCTGACACCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2740
QY 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
DB 2741 CAGGGCAGGGGGGTG-----CCGAATTTGATGCGG-----ATGCTTGGAGAG 2782
QY 294 ThrAspIleGlyLeuGlyTyrProSerGlyAsnGlnGlyTyr 307
DB 2783 ACGTGATTGCGGAGTCCGGGATTTGGGATTCGCAACGCGGCTGG 2825

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2279.84 Seconds
(without alignments)
5684.703 Million cell updates/sec

Title: US-09-985-689A-6
Perfect score: 2250
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVQGFQAFSLAIWN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat_31032004.161807.4168/app query.fasta_1.3498
-DB=EST -OPMT=fastap -SUFFIX=xt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 -CGN_1_1708@runat_31032004.161807.4168 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	239	10.6	1605	13	BQ622771 CC Contig
2	236	10.5	640	12	BJ395336 BJ395336
3	227.5	10.1	4198	11	AK29048 Mus muscu
4	213.5	9.5	532	29	AL452424 T. brucei
5	213.5	9.5	594	12	TA319G10P
6	207.5	9.2	2141	13	BQ393752 BJ393752
7	202.5	9.0	601	12	BQ142519 Contig6 M
8	200	8.9	508	28	BQ387574 BJ387574
9	200	8.9	771	14	CA320325 UI-M-FW0
10	198	8.8	1002	29	CNS06D68
11	190.5	8.5	574	29	TA315H10P
12	189	8.4	633	12	BJ369190
13	189	8.4	716	12	BI750157 F902_1090
14	186	8.3	716	28	BZ893395 HL2_0177
15	185	8.2	665	13	BQ770462 UI-M-F10-
16	183.5	8.2	681	14	CB690041 CBST-54-B
17	182	8.1	675	14	CF727824 UI-M-HB0-
18	182	8.1	4662	11	BC060627 Mus muscu
19	181	8.0	530	29	CNS010FO
20	179.5	8.0	545	13	BU575479 TGESTzyb8
21	178.5	7.9	614	9	AJ273402 AJ273402
22	178	7.9	576	14	CD295943 StrPu691.
23	177.5	7.9	641	12	BJ393925 BJ393925
24	177	7.9	3091	11	BC011275 Mus muscu
25	174	7.7	594	14	CF846026 dBH036XA
26	174	7.7	2121	28	BZ442495 170023066
27	173.5	7.7	650	9	AJ274038
28	173.5	7.7	1029	29	CNS071DW
29	173	7.7	895	13	BQ216158 AGENCOURT
30	171.5	7.6	616	14	CF138007 UI-HF-BN0
31	171.5	7.6	712	14	CD311344 StrPu691.
32	171	7.6	564	28	AQ651427 Sheared D
33	170.5	7.6	573	14	CA937626 sav42b10.
34	169.5	7.5	655	28	BZ349201 hg88C10.9
35	168.5	7.5	580	9	AJ273745
36	168.5	7.5	583	9	AJ273947
37	168.5	7.5	593	9	AJ273918
38	168.5	7.5	601	9	AJ273921
39	167.5	7.4	1572	11	AY107161 Zea mays
40	166.5	7.4	601	9	AJ273050
41	166.5	7.4	601	12	BJ365857
42	166.5	7.4	930	13	EX390734 BX390734
43	166.5	7.4	937	12	BQ246418
44	166	7.4	449	12	BJ359939
45	165.5	7.4	604	9	AJ273185

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

BQ622771 1605 bp mRNA linear EST 01-JUL-2002
CC Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
BQ622771.1 GI:21649940
EST.
Conidiobolus coronatus (Delacroixia coronata)
Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
1 (bases 1 to 1605)

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
TITLE EST analysis of genes expressed by the zygomycete pathogen
Conidiobolus coronatus during optimized secretion of proteins
JOURNAL Unpublished (2002)
COMMENT Contact: Freimoser F. M.
Department of Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA
Tel: 301 405 16 13
Fax: 301 314 92 90
Email: ff34@umd.edu.
FEATURES Location/Qualifiers
source 1..1605
/organism="Conidiobolus coronatus"
/mol_type="mRNA"
/strain="ARSEF 512"
/db_xref="taxon:34488"
/clone_lib="Conidiobolus coronatus ARSEF 512"
/note="Vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."
ORIGIN
Alignment Scores:
Pred. No.: 2,46e-13 Length: 1605
Score: 239.00 Matches: 85
Percent Similarity: 43.99% Conservative: 43
Best Local Similarity: 29.21% Mismatches: 83
Query Match: 10.62% Indels: 80
DB: 13 Gaps: 14
US-09-985-689A-6 (1-434) x BQ622771 (1-1605)
QY 6 GlyIleValIysAlaAspValAlaGlnSerSer-----TyrGlyLeu----- 19
DB 718 GGTGTGTCTAGAGCTATTGCCGTCCAGCCATGCTCTGGGGCTTCTCGTGGTTGGT 777
QY 20 -----TyrGlyGlnGly 23
DB 778 CAACGTGCTAAGCTCGGATCTGCTCTTACTCTTACACCATGCTGATGCTGCAAGGT 837
QY 24 GlnValValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHis 43
DB 838 GTCACTGTTTTCGTTTATGATCTGCTGCTCAATGTCACCCACATGAC----- 885
QY 44 GluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAla--- 62
DB 886 -----TTGGGTGGTGGTGGCCACT-----TGGGGTACTTAACACTGCTGGT 924
QY 63 -----AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
DB 925 GGTAGCAACACTGATGCTCAGCGTCAGGTCCTACTGCTGCTGCTGCTGCTGCTGCTGCT 981
QY 80 GlyThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMet--- 98
DB 982 -----ACCACCTATGTTGTTCCCAAGAGGCTAACATTTGCGGTTAAGTCTTAGT 1035
QY 99 ---AspSerAsnGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGln 117
DB 1036 GATGATGGTCCGGATCACTACTGGAATTATCTCGGTATTGACTGGGTGTTAAG--- 1092
QY 118 AlaTyrSerAlaGlyAlaAlaIleHisThrAsnSerTrpGlyValaProValAsnGlyAla 137
DB 1093 ---CACTCTGCTCCCAAGAAAGTTATCTCTATGAGTTTAGGAGGTGGTAAGAACATGCT 1149
QY 138 TyrThrThrAspSerArgAsnValaPheAspTyrValArgLysAsnAspMetAlaValLeu 157
DB 1150 CTTAACACT-----GCTGTATAACACGCTGTGACAGGGAGTTGCTACTGTTGTC 1200
QY 158 PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
DB 1201 ---GCTGTGGTAAACGATACAGAGATGCTTGTGGATCTCT---CCCGCTTCTGCTCT 1254

QY 178 AsnAlaIleThrValGlyValAlaThrGlu-----AsnLeuArgProSerPheGlySerTyr 195
DB 1255 TCCGCCATTACCGTGTGGTGCACCTGATGATCAATGAATAAAGGGCTTCTTCTTCACTTC 1314
QY 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
DB 1315 GGTAGCTGTGC----- 1326
QY 216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 235
DB 1327 -----GATATCTTAGCTCTCTGGTGTCAACATTCTCTCCACC----- 1362
QY 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 255
DB 1363 -----TGGAGGGATCTTAACACTGCCACCAACACCATCTCTGTTACCTCT 1407
QY 256 MetAlaThrProIleValAlaGlyAsnValAla 266
DB 1408 ATGGCTTGGCTTCACATTGCTGGTTAGCTGCT 1440
RESULT 2
BJ395336 640 bp mRNA linear EST 08-MAR-2002
LOCUS Dictyostelium discoidium cDNA clone dds38b16 5', mRNA sequence.
DEFINITION
ACCESSION BJ395336
VERSION BJ395336.1 GI:19306422
KEYWORDS EST.
SOURCE Dictyostelium discoidium
ORGANISM Dictyostelium discoidium
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 640)
AUTHORS Urushinara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoidium at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1..640
/organism="Dictyostelium discoidium"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds38b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoidium cDNA library, SF"

ORIGIN
Alignment Scores:
Pred. No.: 1.03e-13 Length: 640
Score: 236.00 Matches: 69
Percent Similarity: 49.77% Conservative: 37
Best Local Similarity: 32.39% Mismatches: 69
Query Match: 10.49% Indels: 38
DB: 12 Gaps: 10
US-09-985-689A-6 (1-434) x BJ395336 (1-640)
QY 19 LeuTyrGlyGlnGlyGlnValValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37
DB 22 TTAAGAGGTAAGGTCAAGATATTGAGTATTGCTGATCTGTTTAGATTGTTAGTACCATTTGT 81
QY 38 -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48
DB 82 TTTCTTTTCAGATTCAAAAGTATCCATACCATTTTAATCAAGTGAATGAATCAATAGAAA 141

QY 49 LysileThrAlaileTyrrAlaLeuGlyArThrAsnAsnAlaAspPro-----Asn 66
 Db 142 GTTGTAACTATATT-----ACTTACCATGACAATGAAGATTATGTAAT 186

QY 67 GlyHisGlyThrHisValalaglySerValLeuGlyAsnGly----- 80
 Db 187 GGTCAATGGTACACATGTTTGTGGCTCTGCAGAGGTACTCCAGAGGATCTTCATGGGCT 246

QY 81 ---ThrSerAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAsp 99
 Db 247 ATTTCATCATTTAGTGGTCTGTCACTGACGCAAGATTGCTATTTATGAT---CTTTCA 303

QY 100 SerAsnGlyGlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrr 119
 Db 304 TCTGGAAGTTCTTGAACCAACACCCGAGATACAGTCAATATGATCAAAACCATATAT 363

QY 120 SerAlaGlyAlaAlaArgIleHisThrAsnSerTrpGlyAla-----ProValAsn 135
 Db 364 GATGCGAGGTGCAAGAGTACATGCTGATCTTGGGTTCTGTATCTTTCGCAAGCTTATAT 423

QY 136 GlyAlaTyrrThrThrAspSerArgAsnValAspAspTyrrValArgLys---AsnAspMet 154
 Db 424 GGTGGTATTCCGATGATGCTGGTGTATTTGATGTCATCTCTATGATGATCCAGAAATTC 483

QY 155 AlaValLeuPheAlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGly 174
 Db 484 TCTATCTAAGAGTCTGCTGTAAT---AAGAGCTATTTCATCTTTATAGCTCAAGCA 540

QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
 Db 541 ACAGCTAAAAATGCAATTACAGTTGGTGTCTGACAAACAGCTCATGTAATATATGTGTC 600

QY 195 -----TyrrAlaAspAsnIleAsn 200
 Db 601 GATGCAATGGAATATTGATTTCTCAGATAATGCTAAT 639

RESULT 3
 AK029048
 LOCUS
 DEFINITION
 Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:4732484M11 product:membrane-bound transcription factor protease, site 1, full insert sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 AK029048.1 GI:26325017
 HTG; CAP trapper.
 Mus musculus (house mouse)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 4
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 5
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 6
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 7
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 8
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 9
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 10
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 11
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 12
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 13
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 14
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 15
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 16
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 17
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 18
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 19
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 20
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 21
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 22
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 23
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 24
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 25
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 26
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 27
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 28
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 29
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 30
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 31
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 32
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 33
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 34
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 35
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 36
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 37
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 38
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 39
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 40
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 41
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 42
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 43
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 44
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 45
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 46
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 47
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 48
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 49
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 50
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 51
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 52
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 53
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 54
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 55
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 56
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 57
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 58
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 59
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 60
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 61
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 62
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 63
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 64
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 65
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 66
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 67
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 68
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 69
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 70
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 71
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 72
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 73
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 74
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 75
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 76
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 77
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 78
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 79
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 80
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 81
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 82
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 83
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 84
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 85
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 86
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 87
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 88
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 89
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 90
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 91
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 92
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 93
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 94
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 95
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 96
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 97
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 98
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 99
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 100
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 101
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 102
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 103
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 104
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 105
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 106
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 107
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 108
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 109
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 110
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 111
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 112
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 113
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 114
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 115
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 116
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 117
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 118
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 119
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 120
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 121
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 122
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 123
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 124
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 125
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 126
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 127
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 128
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 129
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 130
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 131
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 132
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 133
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 134
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 135
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 136
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 137
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 138
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 139
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 140
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 141
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 142
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 143
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 144
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 145
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 146
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 147
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 148
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 149
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 150
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 151
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 152
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 153
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 154
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 155
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 156
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 157
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 158
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 159
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 160
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 161
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 162
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 163
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 164
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 165
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 166
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 167
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 168
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 169
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 170
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 171
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 172
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 173
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 174
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 175
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 176
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 177
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 178
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 179
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 180
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 181
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 182
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 183
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 184
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 185
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 186
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 187
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 188
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 189
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 190
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 191
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 192
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 193
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 194
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 195
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 196
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 197
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 198
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 199
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 200
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 201
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 202
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 203
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 204
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 205
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 206
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 207
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 208
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 209
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 210
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 211
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 212
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 213
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 214
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 215
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 216
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 217
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 218
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 219
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 220
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 221
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 222
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 223
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 224
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 225
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 226
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 227
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 228
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 229
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 230
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 231
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 232
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 233
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 234
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 235
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 236
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 237
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 238
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 239
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 240
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 241
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 242
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 243
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 244
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 245
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 246
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 247
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 248
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 249
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 250
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 251
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 252
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 253
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 254
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 255
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 256
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 257
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 258
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 259
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 260
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 261
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 262
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 263
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 264
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 265
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 266
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE

SVTKKAASWEGIAQGHIMITVASPABTELHSGAEHTSVKPIKVIITPPSRKVL
WQYHNLRYPPGYFPDRNLKMDPLDWMGDHVTNFRDMYOHLKSMGYFVEVLGAPF
TCFDAQTGTLIVSEEEYFPEIEAKLRDNDGLSLVIFSDWNTYVMRKVKFYDE
NTROMWPTGGGANIPALNELLSSVNMGFSGLYEGEFVLANDHMYASGCSIAKFFE
DGVILQTFKDOGLEVLKQETAIVENPILGLYQ:PECGGRIVLYGDSNCLDSDHLEA
KDCFWLLDALLQVTSVGVTPBSLSHSGNEORPPSGAGLAPPERMEGNHLRYSKVLEA
HUGDPKPRPULPACFHLUSWAKPQELNETAPSNLWKQKLLSIDLDKVLVLPNPSRNPQV
RPLSPGSGAWDIPGGIMPEGRNVQGVTPVFALGANVAQAFVTVQISKAKSRPKR
RRPRAKRPLQAOQHAPARTFSV
4180..4185
/note="putative"
polyA_signal
4198
/note="putative"
polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 2e-11 Length: 4198
Score: 227.50 Matches: 115
Percent Similarity: 39.03% Conservative: 54
Best Local Similarity: 26.56% Mismatch: 177
Query Match: 10.11% Indels: 87
DB: 11 Gaps: 21

US-09-985-689A-6 (1-434) x AK029048 (1-4198)

Qy 8 VallysAlaaspValAlaGlnSerSerTy:GlyLeuTyGlyGlnGlyGlnValAla 27
Db 1005 CTGAGGAGCATGTCTGTGG---CAGATGGATACACAGGTGCTAATGTCAGATGCT 1061
Qy 28 VallaAlaaspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 1062 GTTTTGATATGGGCTC-----AGTGAGAAGCATCCGCATTTTAAG 1103
Qy 48 GlylysIleThrAlaIleTyAlaLeuGlyArgThrAsnAsnAlaAsn--- 63
Db 1104 AAT-----GTGAGGAGAGACCACTGGACCATTGAGGAGCCCTG 1145
Qy 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSer 82
Db 1146 GATGATGGCTAGGCCATGCGACATCTGTCAGGT---GTGATGGCAGCATGAGGGAG 1202
Qy 83 AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGly 102
Db 1203 TGCCAAAGATTGCTCCAGATGAGAGCTGACATCTTCAGGGTCTTACCAACAT--- 1259
Qy 103 GlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTySerAlaGly 122
Db 1260 -----CAGGTGCTTACACATCTGTTCTGATGCTTCACTATGCC 1304
Qy 123 AlaArgIleHisAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThraSpSer 142
Db 1305 ATCCTAAAGAGATGAGCTTCTCAACCTTAGCATGCTGGGCCGACCTTCATGATCAT 1364
Qy 143 ArgAsnValAspAspTyx-----ValArgLysAsnAspMetAlaValLeuPheAlaAla 160
Db 1365 CGTTTGTGCAAGGTGTGGGAATTAACAGCTAACATGAATATGTTCTGCTATT 1424
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1425 GGCAATGATGAGCTCTCTATGCGCACTCTGAATAACCTGCTGATCAGATGATGATT 1484
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
Db 1485 GGNUGGTTGGCATTCAC-----TTTGAAGATAACATC--- 1517
Qy 201 HisValAlaGlnPheSerSerArgGly-----ProThrLysAspGly 214
Db 1518 -----GCTCGCTTTCTTCCAGGGAATGACTACCTGGGAATTACAGAGGCTATGGT 1571
Qy 215 ArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu 234
Db 1572 CGTGTGAAGCCTGACATGTGC-----ACCTATGGTCTGGAGTGGGGTTCGGT 1622

Qy 235 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThr 254
Db 1623 GTGAAAGGGGGCTGC-----CGTGCACTCTCAGGAC 1655
Qy 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 274
Db 1656 AGTGTGCTTCCCAGTGTGCTGGGCGCTCACCTTGTAGTAAACACAGTACACAG 1715
Qy 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaThr 294
Db 1716 CGGAGAGCTGGTG-----AATCTGCCAGTGTGAGCAAGCTTTGATAGCGTCAGCCGG 1769
Qy 295 AspIleGlyLeuGlyTyProSerGlyAsnGlnGlyTrpGlyArgValThrLeuAspLys 314
Db 1770 AGACTTCTCT---GGGTCACACATGTCGAGCAAGTTCATGGCAAGTTGATGCTGCTCGCA 1826
Qy 315 SerLeuAsnValAlaPheValAsnGluThrSerSerLeuSerThrAsnGlnLysAlaThr 334
Db 1827 GCTTATCAGATC-----CTACGAGCTATAAACCGCAGCAGCCTGAGT 1871
Qy 335 TyrSerPheThrAlaGlnSerGlyLysProLeuLysIleSerLeuValTrp-----Ser 352
Db 1872 CCTAGCTACATGACCTGACTGAGTGTCCC-----TACATGTGCGCTACTGTC 1919
Qy 353 AspAlaProAlaSerThrSerAlaSerValThrLeuValAsnAspLeuAspLeuValIle 372
Db 1920 TCCAGCCTTACTTACTATGAGGAATGCCAACATCGTTAAT-----GTC 1964
Qy 373 ThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyAspAsnAsn 392
Db 1965 ACCATCTCAATGGCATGGGTGCACAGGAAGATTGTGATAAGCCT-----GAG 2015
Qy 393 Trp-----AspGlyArgAsnAsnValGluAsnValPhe----- 403
Db 2016 TGCGACCCCTATTTACACAGATGGAGACACATTGAAGTGGCTTCTCTACTCTCTCA 2075
Qy 404 IleAsnAlaProGlnSerGlyThrTyThrValGluVal 416
Db 2076 GTGTTGTGGCTGTGCTAGTTACCTGCCATCTCCATT 2114

RESULT 4

TA319G10P

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

TA319G10P 532 bp DNA linear GSS 13-DEC-2000
T. brucei sheared genomic DNA clone 319g10, forward sequence,
genomic survey sequence.
AL492464.1 GI:11867408
AL492464.1
GSS.
Trypanosoma brucei
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 532)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.B., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

[illegible]

US-09-985-689A-6 (1-434) x BU387574 (1-601)

```
QY 198 AnNleAAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLys 217
Db 51 AATGAGAAATATATTGTTTCATCAAAAAGGTCCACACATGATGGTAGAATGAAA 110
QY 218 ProAspValMetAlaProGlyThrPheLeuSerAlaAspSerSerLeuAla 235
Db 111 CTGATTAGTTCCTGCTGGTGAATATATTACATCGGCAAGATCAATGGTCCCAATACA 170
QY 236 -----ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMet 251
Db 171 ACAGACCAATGTGGTGGCTCTTTA-----CCAAATACAAATGCATTATTGGCG---ATA 224
QY 252 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 271
Db 225 TCTGGTACATCAATGSCAACCTCATTTGCAGCAGCAGCAGCAACAAATCTTAGACAATAT 284
QY 272 -----PheLeuLysAsnArgGlyIleThrProLys 281
Db 285 TTAGTTGATGGTTATTATCCAACTGGTTCAATTGTAGATCAATAAATAATACAACT 344
QY 282 ProSerLeuLeuLysAlaAlaLeuLysAlaGlyAla----- 293
Db 345 GGATCATTAATAAAGCATTAATGATTAATATGCTCAGTTATTAAATGGTACATTCAA 404
QY 294 -----ThrAspIleGlyLeuGlyTyrProSerGlyAsn----- 304
Db 405 TTGATTACATCATCAAGTATTACATATCCATCAACCAAGTTTTCGAGGT 464
QY 305 -----GlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 320
Db 465 GCAAGTTTATGTTCAAGTTGGGTGCTATTAGAAATGAGTAATGGTTACATGTT----- 518
QY 321 ValAsnGlnThrSerSerLeuSerThrAsnGln 331
Db 519 GTCAATAATAATAATAGTAATAATAATAATA 551
```

RESULT 8
AQ652212/c 508 bp DNA linear GSS 22-JUN-1999
LOCUS Sheared DNA-8P2.TF Sheared DNA Trypanosoma brucei genomic clone
DEFINITION Sheared DNA-8P2, genomic survey sequence.
ACCESSION AQ652212
VERSION AQ652212.1 GI:5145398
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 508)
EL-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
Doneison,J., Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other GSSs: Sheared DNA-8P2.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
7912 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ARCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tcdb/mdb/tbdb/>.
Seq primer: M13-Forward
Class: shotgun.
Location/Qualifiers

FEATURES

source
1..508
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone_lib="Sheared DNA-8P2"
/clone_lib="Sheared DNA"
/notes="Vector: pUC18; Site 1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
Press, 1999)."

ORIGIN

Alignment Scores:
Pred. No.: 3,61e-10 Length: 508
Score: 200.00 Matches: 57
Percent Similarity: 50.89% Conservative: 29
Best Local Similarity: 33.73% Mismatches: 55
Query Match: 8.89% Indels: 28
DB: 7
US-09-985-689A-6 (1-434) x AQ652212 (1-508)
QY 156 ValLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrLysAlaProGlyThr 175
Db 503 CTATCTCTCCATCGACGACAGTATCCAGATGCG-----CTAATGACTCGGTGCGT 450
QY 176 AlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyr 195
Db 449 GGTAAAGACGTGATGCTGGTGGGTGCACAAAAACGTG-----TTGACGCTCG 399
QY 196 AlaAspAsnIleAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArg 215
Db 398 AAGAC-----ATTGTTCTCTGTTTCTTCGATGTCACATACGACGCTAGG 348
QY 216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 235
Db 347 ATGAAACCCGATCTTGTGGTCCGCGGAGAGGTGCTCTCTCTCTCTGCGCAA 288
QY 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 255
Db 287 GCATCAGCAAAA-----CAATGTAAGTGGTGGCCCAAGCGGGTTCTATCG 243
QY 256 MetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys--- 274
Db 242 ATGGCAACTGCGCGCGTCCGCGCGCGCTACACTGTTGCTCAGTATGTACGGAGGCTA 183
QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIle----- 290
Db 182 AATCGAACTGCTTCACTCTGCTGCTTGTCTGAAGCTCTGATGGTCCACTCTACCGGTG 123
QY 291 -----AlaGlyAlaThrAspIleGlyLeuGlyTyr--- 300
Db 122 CCGTTGAGTAATCCACCGTGGAGTGGTTTGGCGCTCTTGATTTATCTCTCTTTTCC 63
QY 301 ProSerGlyAsnGlnGlyTrpGlyArg 309
Db 62 CCAACGGGCACACAGGGATGTTCCCG 36
RESULT 9
CA320325 771 bp. mRNA linear EST 09-JUL-2003
LOCUS UI-M-FWC-cby-d-23-0-UI.r1 NIH_EMAP_FW0 Mus musculus cDNA clone
DEFINITION IMAGE:6816072 5', mRNA sequence.
ACCESSION CA320325
VERSION CA320325.1 GI:24538449

EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 771)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov/>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: PYX-5.
Location/Qualifiers
1. .771
/organism="Mus musculus"
/mol_type="cDNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816072"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP Fw0"
/notes="Organ: Brain; Vector: PYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into PYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGCAGC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN
Alignment Scores:
Pred. No.: 7,42e-10 Length: 771
Score: 200.00 Matches: 83
Percent Similarity: 41.26% Conservative: 35
Best Local Similarity: 29.02% Mismatches: 112
Query Match: 8.89% Indels: 56
DB: 14 Gaps: 12

US-09-985-689A-6 (1-434) x CA320325 (1-771)

Qy 8 VallysAlaSpValAlaGlnSerSerTyrGlyGlnGlyGlnValValAla 27
Dy 40 CTCGAGGCGAGATGTGCTGTGG---CAGATGGGATACACAGGTGCTAATGTCAGAGTTGCT 96
Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Dy 97 GTTTTGTATCTGGGCTC-----AGTGAGAGCATCGGCATTTTAA 138
Qy 48 GlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Dy 139 AAT-----GTGAAGGAGAGAACCACTGGACCAATGACGGACCCCTG 180
Qy 64 ---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSer 82